The Autoregressive Sieve Bootstrap for Random Fields and Multivariate Stochastic Processes

Von der
Carl-Friedrich-Gauß-Fakultät
der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Marco Meyer
geboren am 19. Mai 1983 in Peine

Eingereicht am: 14. November 2014
Disputation am: 17. Dezember 2014

Referent: Prof. Dr. Jens-Peter Kreiß, TU Braunschweig
Koreferentin: Prof. Dr. Anne Leucht, TU Braunschweig
Koreferent: Prof. Dr. Efstathios Paparoditis, University of Cyprus

(2015)

Abstract

In this thesis, we will investigate the range of validity of the autoregressive (AR) sieve bootstrap procedure applied to data from multivariate and spatial processes (the latter are also referred to as 'random fields'). Bootstrap methods are used to approximate the distributions of estimators for finite sample sizes. The AR sieve bootstrap is one particularly attractive method because it merely involves fitting of finite-order autoregressive models which is a well-understood problem. However, we will show that validity of the procedure is not restricted to autoregressive processes, but actually goes far beyond this class.

Multivariate and spatial stationary processes possess an inherent autoregressive structure, as long as mild regularity and smoothness conditions are fulfilled for the spectral density. We will discuss these conditions and develop a theoretical framework that allows us to show that the bootstrap pseudo observations generated by the AR sieve algorithm asymptotically follow this very same autoregressive structure. This property crucially depends on a result called Baxter's inequality, which ensures sufficiently fast convergence of the finite predictors of a process towards its autoregressive coefficients. A version of this inequality for multivariate processes is already available in the literature whereas the generalisation to the case of spatial processes has not been known so far. We will derive such a Baxter-inequality for random fields in this thesis.

We will then derive a criterion which allows us to check for a large class of statistics whether the AR sieve bootstrap works asymptotically. In the very general setting of processes with regular and smooth spectral density, as described previously, we will show that the bootstrap approximation asymptotically does *not* mimic the behaviour of the underlying process, but the one of a slightly modified process. This so-called companion process has many features in common with the underlying process, including all second order properties. The check criterion can then be stated as follows: The AR sieve bootstrap procedure works asymptotically for a specific test statistic if and only if the limiting distributions of the statistic applied to the underlying process on the one hand, and to the companion process on the other hand,

coincide. We will show that this is the case for the sample mean under very mild conditions. However, for sample autocovariances, the procedure can explain why the procedure fails in general. For the sample autocorrelations, the criterion enables us to precisely distinguish between the cases for which the procedure is valid and the ones for which it fails. Moreover, we will show that the AR sieve bootstrap can be used to approximate the distribution of nonparametric trend estimators for multivariate processes.
Finally, we will provide simulation results for sample autocorrelation estimators in the setting of spatial data. It turns out that, for data generated from moving average processes, the AR sieve bootstrap performs very well in comparison to both block bootstrap techniques and normal approximations.

Zusammenfassung

Thema der vorliegenden Arbeit ist die Untersuchung der Validität des Autoregressive-(AR-)Sieve-Bootstrap-Verfahrens für multivariate und räumliche Prozesse (letztere werden auch als ”Zufallsfelder” bezeichnet). Im Allgemeinen werden Bootstrap-Methoden dazu eingesetzt, Verteilungen von Schätzern für endliche Stichprobengrößen zu approximieren. Der AR-Sieve-Bootstrap stellt eine populäre Methode dar, da dieser die Abhängigkeitsstruktur innerhalb der Daten durch das Anpassen autoregressiver Modelle endlicher Ordnung imitiert. Diese Modelle werden in der Praxis vielfach eingesetzt, daher ist die Implementation der Methode unproblematisch. Obwohl dies auf den ersten Blick eine autoregressive Modellannahme impliziert, reicht die Gültigkeit des Verfahrens weit über die Klasse autoregressiver Prozesse hinaus. Tatsächlich besitzen multivariate und räumliche stationäre Prozesse stets eine inhärente AR-Struktur, sofern ihre Spektraldichten regulär und glatt sind. Wir werden eine Reihe von Resultaten bereitstellen, aus denen sich ableiten lässt, dass die vom AR-Sieve-Bootstrap erzeugten Pseudo-Beobachtungen eben diese autoregressive Struktur des zugrunde liegenden Prozesses imitieren. Zu einem wesentlichen Teil basiert dieses Vorgehen auf der sogenannten Baxter-Ungleichung, welche sicherstellt, dass die endlichen Vorhersage-Koeffizienten eines Prozesses unter gewissen Voraussetzungen gegen die beschriebenen autoregressiven Koeffizienten konvergieren. Dieses Resultat ist für multivariate Prozesse bereits bekannt, nicht jedoch für Zufallsfelder, daher werden wir eine solche Baxter-Ungleichung in dieser Arbeit herleiten.
Wir werden im weiteren Verlauf für eine umfangreiche Klasse an Statistiken ein Kriterium entwickeln, welches es dem Anwender erlaubt, mit relativ einfachen Mitteln zu bestimmen, ob der AR-Sieve-Bootstrap für einen bestimmten Schätzer asymptotisch valide ist oder nicht. Unter den beschriebenen, sehr allgemeinen Voraussetzungen – Regularität und Glattheit der Spektraldichte des zugrunde liegenden Prozesses – werden wir beweisen, dass der Bootstrap *nicht* das Verhalten des datengenerierenden Prozesses imitiert, sondern vielmehr das eines modifizierten Prozesses. Dieser sogenannte Companion-Prozess hat allerdings viele Eigenschaften mit dem

zugrundeliegenden Prozess gemein, z.B. alle Merkmale zweiter Ordnung wie Autokovarianzen oder die Spektraldichte. Das Entscheidungskriterium wird schließlich folgendermaßen lauten: Das AR-Sieve-Bootstrap-Verfahren ist asymptotisch konsistent für eine bestimmte Statistik genau dann, wenn die Grenzverteilungen der Statistik basierend auf dem zugrundeliegenden Prozess einerseits, und basierend auf dem Companion-Prozess andererseits, übereinstimmen. Dies ist z.B. unter sehr schwachen Voraussetzungen für das arithmetische Mittel der Fall. Für die Stichprobenautokovarianzen hingegen kann mithilfe unserer Methodik erklärt werden, warum das Verfahren im Allgemeinen nicht funktioniert. Im Falle der Stichprobenautokorrelationen wird sich herausstellen, dass der AR-Sieve-Bootstrap nur unter bestimmten weiteren Voraussetzungen funktionieren kann. Das entwickelte Entscheidungskriterium wird es uns erlauben, diese benötigten Voraussetzungen sehr präzise abzugrenzen. Darüber hinaus werden wir eine Möglichkeit aufzeigen, wie das Verfahren dazu genutzt werden kann, die Verteilungen nichtparametrischer Trendschätzer für multivariate Prozesse zu approximieren.
Im Rahmen einer Simulationsstudie werden wir schließlich die Methode auf Autokorrelationsschätzer im räumlichen Fall anwenden. Die Ergebnisse zeigen, dass der AR-Sieve-Bootstrap sehr gut abschneidet, sowohl im Vergleich mit Block-Bootstrap-Methoden als auch mit Normal-Approximationen.

Contents

List of Figures

1 Introduction

Over the past decades, time series analysis has evolved as one of the most intensively studied subjects in statistics. In many applications, data are collected in the course of time, where the realisation at one time point does have an influence on realisations at other time points. Examples can be taken from nearly every scientific field which uses statistical methods, whether it may be economics, where stock prices or interest rates are observed over time, or meteorology, where temperature or precipitation data are collected throughout the year, to name just a few. In both of the aforementioned examples, the data depend on each other which makes the application of classical statistical methods difficult because these methods often require independent or at least uncorrelated data. Therefore, time series analysts are challenged with the problem to derive models and to impose conditions on the data which are, on the one hand, sufficiently strong to infer useful results and, on the other hand, general enough to cover the wide field of possible applications.

It is usually assumed that the observed time series $X_1, \ldots, X_n$ is generated by a stochastic process $(X_t)_{t\in\mathbb{Z}}$ which is in most cases assumed to be real-valued but which may also be integer- or complex-valued, or even have functional values. (X_t) is often referred to as the *underlying process* or *data-generating process*, and the goal is to gather as much information as possible about this process from a given sample $X_1, \ldots, X_n$. The probably most common assumption in time series analysis is stationarity. A real-valued stochastic process $(X_t)_{t\in\mathbb{Z}}$ is called *strictly stationary* if its finite-dimensional joint distributions are invariant under time shifts, i.e. if

$$\mathcal{L}(X_{t_1}, \ldots, X_{t_d}) = \mathcal{L}(X_{t_1+h}, \ldots, X_{t_d+h}) \quad \forall\, d \in \mathbb{N},\ (t_1, \ldots, t_d) \in \mathbb{Z}^d,\ h \in \mathbb{Z}.$$

If the process has finite second moments, strict stationarity directly implies that there exist $\mu \in \mathbb{R}$ and a function $\gamma : \mathbb{Z} \to \mathbb{R}$ such that

$$E(X_t) = \mu, \quad \operatorname{Cov}(X_{t+h}, X_t) = \gamma(h), \quad \forall\, t, h \in \mathbb{Z}. \tag{1.1}$$

In particular, this means that the covariance structure of the process depends only on the time lag between observations but not on the exact location. A process fulfilling condition (1.1) is called *weakly stationary*, or simply *stationary*, and μ is called the *mean* and γ the *autocovariance function* of the process. The simplest example of a stationary process is a *white noise process*, meaning a process with finite second moments fulfilling (1.1) with $\mu = 0$, $\gamma(0) = \sigma^2 > 0$ and $\gamma(h) = 0$ for all $h \neq 0$.

Stationarity is a particularly useful condition in time series analysis because it provides a basis for statistical inference via empirical means over time. Also, it is in many cases a justified assumption. After removing trends (such as, for example, global warming for temperature data) and seasonalities, many observed time series show stationary behaviour.

A much more restrictive assumption on the underlying process is the linearity condition which, however, allows for considerably stronger results of statistical inference. A stochastic process $(X_t)_{t\in\mathbb{Z}}$ is said to be a *linear process* with mean $\mu \in \mathbb{R}$, if there exist an independent and identically distributed (i.i.d.) white noise process $(e_t)_{t\in\mathbb{Z}}$ and an absolutely summable sequence of coefficients $(b_k)_{k\in\mathbb{Z}}$ such that

$$X_t = \mu + \sum_{k=-\infty}^{\infty} b_k \, e_{t-k}.$$

The stationarity conditions and linear process representation of the underlying process introduced so far are part of the so-called *time domain* analysis of time series. A different approach to characterising properties of stationary processes is the so-called *frequency domain* approach. For a real-valued stationary processes $(X_t)_{t\in\mathbb{Z}}$, with the additional condition $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, the sequence of autocovariances can be interpreted as the sequence of Fourier coefficients of an $\mathbb{R}_{\geq 0}$-valued 2π-periodic function f on $\mathbb{R}$. The absolutely convergent Fourier series

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \, e^{-ih\lambda}, \quad \lambda \in (-\pi, \pi],$$

is then called the *spectral density* of the process (X_t). The advantage of the frequency domain approach is that it allows for using the well-established and comprehensive theory of Fourier transformations to the field of time series analysis. Inference in the frequency domain provides important information about the structure of the underlying process. In fact, as will be discussed later on in this thesis, smoothness properties of the spectral density are sufficient for the process to have autoregressive and moving average representations, which are very useful tools for further inference in the time domain.

1.1 Multivariate and spatial data

In this thesis, two ways of generalising time series analysis will play an important role. Namely, inference for multivariate data on the one hand, and for spatial data on the other hand will be considered. When observing time series in practice, it is natural to observe not only one but multiple quantities – which may depend on one another – at each time point. For example, weather stations will typically measure temperature data, but also wind speed, precipitation, humidity, and so on; and of course, these variables will in general be interdependent. Therefore, it is natural to assume that a multivariate, say $\mathbb{R}^q$-valued, time series $\underline{X}_1, \ldots, \underline{X}_n$ is observed. The theory for univariate stochastic processes introduced so far generalises quite naturally to the multivariate setting. An $\mathbb{R}^q$-valued stochastic process $(\underline{X}_t)_{t\in\mathbb{Z}}$ with finite second moments is called (weakly) stationary if there exist a mean vector $\underline{\mu} \in \mathbb{R}^q$ and a matrix-valued autocovariance function $\Gamma : \mathbb{Z} \to \mathbb{R}^{q\times q}$ such that

$$E(\underline{X}_t) = \underline{\mu}, \quad \operatorname{Cov}(\underline{X}_{t+h}, \underline{X}_t) = E\Big((\underline{X}_{t+h} - \underline{\mu})(\underline{X}_t - \underline{\mu})^T\Big) = \Gamma(h), \quad \forall\, t, h \in \mathbb{Z}.$$

In this context, a process $(\underline{X}_t)$ is called linear if there exists a q-variate i.i.d. white noise process $(\underline{e}_t)_{t\in\mathbb{Z}}$, that is a process with $E(\underline{e}_t) = \underline{0}$, as well as a sequence $(B_k)_{k\in\mathbb{Z}}$ of $\mathbb{R}^{q\times q}$-valued matrices with $\sum_{k=-\infty}^{\infty} \|B_k\| < \infty$, where $\|\cdot\|$ denotes any matrix norm, such that

$$\underline{X}_t = \underline{\mu} + \sum_{k=-\infty}^{\infty} B_k\, \underline{e}_{t-k}.$$

Spectral theory also carries over naturally from the univariate setting: If for all $1 \le i, j \le q$ the (i,j)-th entries of the autocovariance matrices of a stationary process $(\underline{X}_t)$ are absolutely summable, i.e. $\sum_{h=-\infty}^{\infty} |\Gamma^{(i,j)}(h)| < \infty$, then the $\mathbb{R}^{q\times q}$-valued matrix given by the absolutely convergent Fourier series

$$W(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma(h)\, e^{-ih\lambda}, \quad \lambda \in (-\pi, \pi],$$

is called the *spectral density (matrix)* of the process.

A completely different field of statistics for dependent data is opened up if one allows for the underlying 'time'-parameter set to be of higher dimensions. Classical time series analysis considers random variables X_t with $t \in \mathbb{Z}$, and the integer numbers $\mathbb{Z}$ form, in a natural way, a totally ordered set; that is, given any $t_0 \in \mathbb{Z}$, one can naturally distinguish between the 'past' values $t < t_0$ and the 'future' values $t > t_0$ in relation to t_0. Now assume that dependent random variables are observed at certain locations in space. To pick up a previously stated example, one might consider a network of weather stations measuring meteorological data at different locations. These locations are discrete, and can therefore be identified with coordinate vectors $\underline{t} \in \mathbb{Z}^d$. Notice that $\mathbb{Z}^d$ with $d \geq 2$ is no longer totally ordered in a natural way, and there are various possibilities to define an order relation. Observations taken at the locations can be modelled as random variables $\{X_{\underline{t}} : \underline{t} \in T\}$, where T is a finite subset of $\mathbb{Z}^d$. Such data sets are usually described as *spatial data* because the dependence structure of $X_{\underline{t}_1}$ and $X_{\underline{t}_2}$ is determined by the spatial location of $\underline{t}_1$ and $\underline{t}_2$ within $\mathbb{Z}^d$. It is assumed that the data $\{X_{\underline{t}} : \underline{t} \in T\}$ are generated by a stochastic process $(X_{\underline{t}})_{\underline{t} \in \mathbb{Z}^d}$ which is called a *(discrete) spatial process* or, by many authors, a *random field*.

In many applications with spatial observations, after removing trends, the dependence structure of random variables $X_{\underline{t}_1}$ and $X_{\underline{t}_2}$ indeed depends only on the relative location of $\underline{t}_1$ and $\underline{t}_2$, instead of their absolute positions within $\mathbb{Z}^d$. This leads directly to the concept of stationarity for spatial data: As a natural generalisation to the time series case, a real-valued spatial stochastic process $(X_{\underline{t}})_{\underline{t} \in \mathbb{Z}^d}$ with finite second moments is called (weakly) stationary if there exists a mean $\mu \in \mathbb{R}$ and an autocovariance function $\gamma : \mathbb{Z}^d \to \mathbb{R}$ such that

$$E(X_{\underline{t}}) = \mu, \quad \operatorname{Cov}(X_{\underline{t}+\underline{h}}, X_{\underline{t}}) = \gamma(\underline{h}), \quad \forall\, \underline{t}, \underline{h} \in \mathbb{Z}^d.$$

White-noise processes are also defined analogously to the time series case, and $(X_{\underline{t}})$ is called a *linear spatial process* if there exist $\mu \in \mathbb{R}$, an i.i.d. white noise $(e_{\underline{t}})_{\underline{t} \in \mathbb{Z}^d}$ and a d-fold sequence of coefficients $(b_{\underline{k}})_{\underline{k} \in \mathbb{Z}^d}$ with $\sum_{\underline{k} \in \mathbb{Z}^d} |b_{\underline{k}}| < \infty$ such that

$$X_{\underline{t}} = \mu + \sum_{k \in \mathbb{Z}^d} b_{\underline{k}}\, e_{\underline{t}-\underline{k}}.$$

Interestingly enough, although the autocovariance function γ is in this context defined on $\mathbb{Z}^d$, the rich theory of Fourier analysis can still be applied to gain information in the spectral domain. Under the condition $\sum_{\underline{h} \in \mathbb{Z}^d} |\gamma(\underline{h})| < \infty$, the function f defined on $(-\pi, \pi]^d$, given by the absolutely convergent Fourier series

$$f(\underline{\lambda}) = \frac{1}{(2\pi)^d} \sum_{\underline{h} \in \mathbb{Z}^d} \gamma(\underline{h})\, e^{-i\langle \underline{h}, \underline{\lambda} \rangle}, \quad \underline{\lambda} \in (-\pi, \pi]^d,$$

is $\mathbb{R}_{\geq 0}$-valued due to $\gamma(-\underline{h}) = \gamma(\underline{h})$, and is called the *spectral density* of the spatial process $(X_{\underline{t}})$. Here, $\langle \cdot, \cdot \rangle$ denotes the usual (Euclidean) scalar product on $\mathbb{R}^d$.

Chapter 2 of this thesis will be concerned with bootstrap methodology for spatial data, while chapter 3 will deal with multivariate data. Of course, a natural extension would be given by multivariate spatial processes. However, it will be explained in more detail during the following sections why the two aforementioned ways of generalisation are treated separately.

1.2 Overview of bootstrap methods

Gaining information about the distribution of estimators is one of the core tasks of statisticians. In many situations the asymptotic distribution of an estimator of interest is known. However, one is usually interested in the estimator's exact distribution for finite sample sizes, which is often times either hard or impossible to derive analytically. A commonly used approach is to approximate the finite sample distribution by an estimation of the limiting distribution, but this method has some considerable drawbacks: First of all, for many estimators, the limiting distribution is not known and the method is therefore not applicable at all. In other situations, the asymptotic distribution may be known but an estimation may be difficult or impossible to carry out. Moreover, even if the method is in principle applicable, the limiting distribution may differ considerably from the finite sample distribution one is interested in. In particular, many estimators are asymptotically normal, and by fitting a normal distribution – which is, of course, symmetric – one is not able to depict potential skewness of the finite sample distribution appropriately.

Because of these drawbacks, resampling methods and, particularly, bootstrap procedures have become very popular among statisticians. The classical bootstrap concept was introduced by Efron (1979) and works as follows: Suppose a sample of i.i.d. random variables $X_1, \ldots, X_n$, $n \in \mathbb{N}$, is given and we are interested in the distribution of an estimator $T_n = T_n(X_1, \ldots, X_n)$. The following steps are performed:

(1) Generate pseudo observations $X_1^*, \ldots, X_n^*$ by drawing n times with replacement from the set of values $\{X_1, \ldots, X_n\}$.

(2) Calculate the plug-in-estimator $T_{n,(1)}^* = T_n(X_1^*, \ldots, X_n^*)$.

(3) Repeat steps (1) and (2) M times, where M is sufficiently large, in order to obtain independent realisations $T^*_{n,(1)}, \ldots, T^*_{n,(M)}$ of the plug-in estimator.

(4) Use the empirical distribution of $T^*_{n,(1)}, \ldots, T^*_{n,(M)}$ as an approximation for the distribution of T_n.

In many situtions the bootstrap method is shown to be asymptotically valid, i.e. the distribution of T_n and the empirical distribution of $T^*_{n,(1)}, \ldots, T^*_{n,(M)}$ are shown to converge (the latter in probability) towards a common limiting distribution. Although asymptotic validity does not imply that the bootstrap approximation is closer to the true distribution than the approximation by the limit, this desirable feature often times can be observed in practice.

Unfortunately, this classical bootstrap scheme by Efron works only for i.i.d. samples, but not for dependent data as occur in time series analysis. Therefore, more advanced bootstrap techniques have been developed which can deal with dependence in the data, at least under appropriate conditions. Many of these techniques can be attributed to one of the following prominent classes:

Residual bootstrap: These procedures depend on the assumption that the underlying process can be described sufficiently well by a parametric model, such as a finite-order autoregression with (approximately) i.i.d. innovations, for example. In this case, the model parameters are estimated in the first step, and the classical bootstrap concept is then applied to the estimated residuals of the model fit. In a final step, the resampled residuals, along with the fitted model, are used to generate bootstrap replicates of the original data. While this method usually shows very good behaviour as long as the parametric assumption is fulfilled, it may fail completely if the underlying process cannot be described sufficiently well by the chosen parametric model. For more information, see, among others, Freedman (1984), Kreiss (1988).

Block bootstrap: This approach is rather straightforward. The data sample is divided into b blocks, each consisting of l successive observations. The resampling is then carried out for this set of blocks, i.e. b blocks are drawn with replacement and then stuck together to form a bootstrap sample consisting of n observations. In the case of the so-called moving block bootstrap, the b blocks are drawn with replacement from the set of all blocks of l successive observations from the original sample, i.e. from a set of $n - l + 1$ possible blocks. To obtain bootstrap validity, it is assumed that both b and l tend to infinity for $n \to \infty$. The main advantage of the block bootstrap is that it works in a very general setting since no parametric assumption is needed. However, its performance is sensitive with respect to the choice of the block length l. Moreover, in situations where the data can be described well by a parametric model, the block bootstrap is often times outperformed by the suitable parametric bootstrap. For more information, see, among others, Carlstein (1986), Künsch (1989).

Frequency domain bootstrap: These methods are based on periodogram estimators $I_n(\lambda)$ of the spectral density $f(\lambda)$, which fulfil the desirable feature to be asymptotically independent at distinct frequencies. The major drawback of this technique is that it works only for statistics which can be expressed as functions of the periodogram, but the procedure is not able to generate bootstrap replicates in the time domain, i.e. in terms of the original data sample. For more information, see, among others, Franke and Härdle (1992), Dahlhaus and Janas (1996), Kreiss and Paparoditis (2003).

1.3 The autoregressive sieve bootstrap

This thesis centers on the so-called *autoregressive (AR) sieve bootstrap* which can be attributed to the class of residual bootstraps. In contrast to the classical autoregressive bootstrap, which fits an AR model of finite order p to the given data, the AR sieve bootstrap allows for the model order p to depend on the sample size n, and it is assumed that $p(n) \to \infty$, as n tends to infinity. Hence, this procedure works under appropriate conditions for AR(∞)-processes which form a much wider, nonparametric class of processes than the one given by fixed-order autoregressive models. This bootstrap concept was originally developed by Kreiss (1988, 1992, 1997) for linear AR(∞)-processes, while Paparoditis (1996) established its validity for multivariate linear processes. Bühlmann (1997) extended the validity from the class of linear processes with exponentially decaying coefficients to the one of linear processes with polynomially decaying coefficients, and introduced the name autoregressive sieve bootstrap, which has since become the generally accepted term for the procedure in the literature.

The procedure can be described as follows. Let a time series $X_1, \ldots, X_n$ and an estimator $T_n = T_n(X_1, \dots, X_n)$ be given. Instead of step 1 of the classical bootstrap procedure introduced in the previous section, perform the following steps:

(1a) Select an order $p = p(n) \in \mathbb{N},\ p \ll n$ and fit a p-th order autoregressive model to the given observations $X_1, \dots, X_n$, for example via Yule-Walker estimation. Denote by $\widehat{a}_1(p), \dots, \widehat{a}_p(p)$ the estimators of the autoregressive parameters.

(1b) Let $\varepsilon'_t = X_t - \sum_{j=1}^{p} \widehat{a}_j(p) X_{t-j},\ \ t = p+1, \dots, n$, be the residuals of the autoregressive fit and denote the centered residuals by $\widehat{\varepsilon}_t = \varepsilon'_t - \overline{\varepsilon}$, where $\overline{\varepsilon} = (n-p)^{-1} \sum_{t=p+1}^{n} \varepsilon'_t$. Here, we suppress the dependence of ε'_t and $\widehat{\varepsilon}_t$ on p for convenience reasons. Generate a sufficient number of independent random variables $\varepsilon_1^*, \varepsilon_2^*, \dots$ by drawing with replacement from the set of centered residuals. Use these bootstrap innovations ε_t^* and the fitted model to calculate a bootstrap sample $(X_1^*, \dots, X_n^*)$ according to the generating equation

$$X_t^* = \sum_{j=1}^{p} \widehat{a}_j(p) X_{t-j}^* + \varepsilon_t^*.$$

Then proceed with steps 2 to 4 of the classical bootstrap procedure from section 1.2.

Asymptotic validity of the AR sieve bootstrap for linear processes was established in Kreiss (1988), Kreiss (1997) and Bühlmann (1997). However, Kreiss, Paparoditis and Politis (2011) discovered that the procedure, at least for certain statistics, is actually valid for a much wider class of processes. Namely, this is the class of processes with smooth spectral densities which are bounded away from zero since these processes possess autoregressive representations

$$X_t = \sum_{j=1}^{\infty} a_j X_{t-j} + \varepsilon_t, \tag{1.2}$$

with suitable coefficients $(a_j)_{j \in \mathbb{N}}$ and with an uncorrelated (but not i.i.d.) white noise process $(\varepsilon_t)_{t \in \mathbb{Z}}$. (ε_t) is called the innovation process, for a more detailed explanation see section 1.4. In addition, Kreiss et al. introduced the so-called *companion process*, a modified version of the underlying process that serves as a comparative quantity and that yields a check criterion for bootstrap validity. To be precise, assume a strictly stationary process $(X_t)_{t \in \mathbb{Z}}$ with representation (1.2) is given. Then, $(\varepsilon_t)_{t \in \mathbb{Z}}$ is strictly stationary as well, and one can define the i.i.d. process $(\widetilde{\varepsilon}_t)_{t \in \mathbb{Z}}$ with marginal distribution $\mathcal{L}(\widetilde{\varepsilon}_t) = \mathcal{L}(\varepsilon_t)$. The companion process is then given by

$$\widetilde{X}_t = \sum_{j=1}^{\infty} a_j \widetilde{X}_{t-j} + \widetilde{\varepsilon}_t,$$

with the very same autoregressive coefficients as in (1.2). Note that the only difference between (X_t) and $(\widetilde{X}_t)$ is the dependence structure of the innovations, while all second order properties, such as autocovariances and spectral densities, coincide. Kreiss, Paparoditis and Politis (2011) showed for a large class of statistics that the AR sieve bootstrap asymptotically mimics the behaviour of the companion process $(\widetilde{X}_t)$ instead of the one of (X_t). Hence, the check criterion can be stated as follows: The AR sieve bootstrap procedure asymptotically works for a process (X_t) and an estimator T_n, if and only if the limiting distributions of T_n applied to (X_t) and T_n applied to $(\widetilde{X}_t)$ coincide. For example, this is the case whenever the limiting distributions depend only on autocovariances/autocorrelations or the spectral density of the underlying processes.

It will be the main purpose of this thesis to derive similar results for the case of multivariate and spatial data. As was already mentioned in section 1.1, one could in principle look at the most general setting of an $\mathbb{R}^q$-valued random field with time domain $\mathbb{Z}^d$. However, although the final results are quite similar, it is interesting that on the way of generalising the theory to multivariate settings on the one hand, and spatial settings on the other hand, we encountered completely different kinds of problems.

For multivariate data, as expected, one has to deal with the problem that simple multiplication of coefficients or autocovariances becomes multiplication of matrices, which is not commutative, leading to various difficulties in the calculations. Things are further complicated by the fact that, in order to establish stability of fitted autoregressive models, we have to consider roots of the determinants of matrix-valued polynomials instead of complex-valued polynomials in the time series case. On a positive note, Baxter's inequality – a key tool for proving validity of the AR sieve bootstrap – is already available in the multivariate setting; for more information see section 1.4.

For spatial data the situation is completely different: Here, the time domain $\mathbb{Z}^d$ is no longer completely ordered in a natural way, which brings up the question which kind of 'one-sided' autoregression, similar to (1.2), can be derived for the random field setting in the first place. Many necessary properties of the autoregressive representations will be established in chapter 2. These derivations deal with problems which are inherent to the spatial structure of the 'time'-domain and should not be confused with the aforementioned difficulties for multivariate processes. Furthermore, in contrast to the multivariate time series setting, a generalisation of the important Baxter-inequality to random fields has not been established up to this point. We will derive such an inequality for real-valued spatial processes in section 2.2.1. However, we conjecture that neither our proof nor Baxter's original proof for the time series case can be generalised to the setting of multivariate random fields. The following section explains this problem in more detail.

Because of the reasons mentioned above, we will treat the multivariate and the spatial setting separately throughout this thesis. Chapter 2 is thus concerned with the generalisation to the case of random fields, while chapter 3 deals with the multivariate situation.

1.4 Baxter's inequality

One of the key results for proving asymptotic validity of the autoregressive sieve bootstrap is Baxter's inequality, derived by Baxter (1962). In short words, it connects the finite-order prediction coefficients of a stationary process with its infinite-order autoregressive coefficients. To elaborate this, let a real-valued stationary process $(X_t)_{t\in\mathbb{Z}}$ be given which possesses a spectral density f bounded away from zero, and autocovariances fulfilling $\sum_{h\in\mathbb{Z}} \nu(|h|)\,|\gamma(h)| < \infty$ for a suitable weight function ν. The latter condition merely ensures that the autocovariances decay with an appropriate rate. Then there exists a sequence of coefficients $(a_j)_{j\in\mathbb{N}}$ which fulfil $\sum_{j\in\mathbb{N}} \nu(j)\,|a_j| < \infty$, such that $\widehat{X}_t$, the L^2-projection of X_t on its infinite past, has the representation

$$\widehat{X}_t = \sum_{j=1}^{\infty} a_j X_{t-j}.$$

Here, L^2-projection means that $\varepsilon_t := X_t - \widehat{X}_t$ is orthogonal to each X_{t-j}, $j \in \mathbb{N}$, i.e. $\mathrm{Cov}(X_{t-j}, \varepsilon_t) = 0$. $(\varepsilon_t)_{t\in\mathbb{Z}}$ is then called the *innovation process* of (X_t). Obviously, such a process possesses the representation (1.2).

Now let $a_1(p), \ldots, a_p(p)$ be the L^2-projection coefficients of X_t with respect to its finite past $X_{t-1}, \ldots, X_{t-p}$, i.e. let $\mathrm{Cov}(X_t - \sum_{j=1}^{p} a_j(p)\, X_{t-j}, X_{t-k}) = 0$ be fulfilled for all $k = 1, \dots, p$. Obviously, these coefficients fulfil the Yule-Walker equations

$$\begin{pmatrix} \gamma(0) & \cdots & \gamma(p-1) \\ \vdots & \ddots & \vdots \\ \gamma(p-1) & \cdots & \gamma(0) \end{pmatrix} \begin{pmatrix} a_1(p) \\ \vdots \\ a_p(p) \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(p) \end{pmatrix}. \tag{1.3}$$

When fitting a p-th order autoregressive model to a data sample $X_1, \dots, X_n$, as is required in step ($1a$) of the AR sieve bootstrap procedure, one often uses Yule-Walker estimators, which can be obtained from replacing the autocovariance function $\gamma(\cdot)$ by its empirical version $\widehat{\gamma}(\cdot)$ in (1.3) and then solving the linear system. By an appropriate choice of p one can ensure that the coefficients $a_1(p), \dots, a_p(p)$ are estimated consistently. Hence, it arises the question whether the projection coefficients $a_1(p), \dots, a_p(p)$ actually converge towards the AR coefficients $(a_j)_{j \in \mathbb{N}}$ for $p \to \infty$, and, if so, how the rate of convergence can be specified.

In the aforementioned paper Glen Baxter established the following result, called *Baxter's inequality*: As long as the weight function fulfils $\nu(k) \leq \nu(j)\,\nu(|k-j|)$ for all $k, j \in \mathbb{N}$, there exists $C < \infty$ and $p_0 \in \mathbb{N}$ such that

$$\sum_{j=1}^{p} \nu(j)\,|a_j(p) - a_j| \leq C \cdot \sum_{j=p+1}^{\infty} \nu(j)\,|a_j|, \qquad \forall\, p \geq p_0. \tag{1.4}$$

Due to the summability property of $(a_j)_{j \in \mathbb{N}}$, the right-hand side converges to zero, as $p \to \infty$. By choosing an appropriate weight function ν, this inequality yields the desired rate of uniform convergence for each $a_j(p)$ towards a_j.

Since we will be extending the AR sieve bootstrap concept to multivariate and spatial data, we will also need Baxter-type inequalities in these settings. For the multivariate setting, it is known that the inequality holds just as in the time series case, merely replacing coefficients by coefficient matrices and $|\cdot|$ by a matrix norm $\|\cdot\|$, cf. Hannan and Deistler (1988). However, in the spatial setting, a result in the spirit of (1.4) has not been developed so far. This might be due to the fact that Baxter's original proof relies in large parts on a result concerning Fourier coefficients of complex-valued polynomials, cf. Theorem 1.1 in Baxter (1963). We conjecture that this result cannot be extended to the case of Fourier series in several variables which makes a generalisation of the original proof of Baxter's inequality to random fields impossible. However, we will establish a Baxter-type inequality in section 2.2.1 of this thesis, using a completely different strategy. Our result will not be as general as Baxter's result for time series, but instead hold for a specific class of weight functions, only. Still, this inequality will be sufficient for establishing the rate of convergence of finite predictor coefficients that is required for proving validity of the AR sieve bootstrap.

The following consideration is connected to Baxter's inequality but might be of its own interest: In the time series case, as mentioned earlier, it is well-known that the weighted summability property of the autocovariances carries over to the autoregressive coefficients, i.e. that $\sum_{j\in\mathbb{N}} \nu(j)\,|a_j| < \infty$, which essentially yields a rate of decay for the AR coefficients. This result can be deduced from a weighted version of Wiener's Lemma, sometimes also referred to as the first assertion of the Wiener-Lévy-Theorem, cf. Zygmund (2002), Chapter VI, Theorem 5.2. Of course, Baxter's inequality (1.4) would be useless without this property. The problem with our approach for the random field case is the fact that, for the weight functions we will be considering, a similar result about the rate of decay of the AR coefficients seems not to be available in the literature.

We will therefore provide such a result in section 2.1. To be precise, we will first invoke a weighted version of Wiener's Lemma in several variables to derive weighted summability of the Fourier coefficients of the logarithmised spectral density, the so-called cepstral coefficients of the process. Then we will carry over the established weighted summability of the cepstral coefficients to the AR coefficients. This result provides a rate of decay for the autoregressive coefficients that is purely based on properties of the spectral density and may be of its own interest.

1.5 Acknowledgements

First and foremost, I would like to thank my advisor Jens-Peter Kreiß for offering me a position in his group, and for his constant support on my way of preparing this thesis. He always encouraged me and placed trust in my abilities, whether it was research or teaching. In particular, I appreciate him giving me the opportunity to deal with mathematical problems at my own pace, without applying any pressure; yet for always being available for extensive discussions and offering a helping hand whenever needed.

I also thank Carsten Jentsch for endless hours of working together on the generalisation of the original proof of Baxter's inequality which, unfortunately, could never be completed. However, the much more straightforward proof of the slightly weaker version of the result provided in this thesis would probably still not be completed without his input.

My gratitude also goes to Anne Leucht and Efstathios Paparoditis, who immediately agreed to serving as referees of this thesis.

Furthermore, I would like to thank all current and former members of the Institute for Mathematical Stochastics in Braunschweig, for providing a pleasant working atmosphere and a culture of helping each other out.

Last but not least, I would like to thank Laura and my parents for always encouraging and supporting me on my way.

2 The Autoregressive Sieve Bootstrap for Random Fields

Based on: Meyer, M., Jentsch, C. and Kreiss, J.-P.:
Baxter's Inequality and Sieve Bootstrap for Random Fields.
Preprint (2014).

In this chapter we will generalise the concept of the autoregressive (AR) sieve bootstrap to the case of real-valued stationary random fields $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$, which are also often called spatial processes. The entire theory that will be established can be generalised to spatial processes in higher spatial dimensions, i.e. $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^d}$, $d \in \mathbb{N}$, in a straightforward way. However, for the sake of notational convenience, we will restrict ourselves to the case $d = 2$ here, since all major differences between the time series case ($d = 1$) and the random field case ($d \geq 2$) can be discussed thoroughly for $d = 2$. Characterising conditions for asymptotic validity of the bootstrap procedure will be done in the spirit of Kreiss, Paparoditis and Politis (2011), who explored the range of validity for the time series case, see also section 1.3 for a short summary.

This chapter is organised as follows: In section 2.1 we will introduce the basic notations and definitions and formulate the algorithm of the AR sieve bootstrap procedure precisely. Moreover, we will start our discussion with the fundamental question what kind of counterpart one-sided autoregressive representations, as given by (1.2) for the time series case, have in the spatial setting. Whittle (1954) discovered that stationary spatial processes, under very mild conditions on the spectral density, possess AR representations with respect to half-planes of $\mathbb{Z}^2$ which are suitable for our purposes; this will be further elaborated in section 2.1. We will also clarify a common misunderstanding in the discussion of spatial and time series autoregressions, that should at least be mentioned briefly at this point: It is often criticized that, for spatial processes, one has to choose a concept of 'past' values for one-sided autoregressions, i.e. choose a direction from which the random variables $X_{\underline{t}}$ are influenced.

This choice is of course arbitrary. Hence, one might come to the conclusion that the whole concept of one-sided autoregressions implies a very specific model assumption which is not fulfilled for real-world data. However, the opposite is true since our assumptions do not constrain the class of processes any further than demanding the spectral density to be positive and smooth. This point will be discussed in section 2.1.

While Whittle (1954) established the aforementioned AR representations with absolutely summable autoregressive coefficients, he did not specify conditions to ensure polynomial rates of decay for the coefficients, which are needed to obtain validity of the AR sieve bootstrap. Therefore, we will derive the required rates in section 2.1. To be precise, we will connect the AR coefficients to the Fourier coefficients of the logarithmised spectral density, the so-called cepstral coefficients, and show how the required rate of decay carries over from the autocovariances of the underlying process to the cepstral coefficients and then to the AR coefficients.

In section 2.2, we will establish sufficiently fast convergence of the finite-order AR models that are fitted in the course of the sieve bootstrap procedure, to the aforementioned AR coefficients. Here, a key result is a generalisation of Baxter's inequality, cf. Baxter (1962) and also section 1.4, to the case of random fields. Beyond its application in connection with the AR sieve bootstrap, this result may be of its own interest.

The exact conditions for AR sieve bootstrap validity are given in section 2.3, and the result will be a check-criterion which allows to decide whether the procedure works asymptotically or not; with the criterion being solely based on the asymptotics of the particular test statistic one is looking at. This result closely resembles the concept of the so-called companion process, introduced by Kreiss, Paparoditis and Politis (2011), see also section 1.3. We will apply the derived check criterion in section 2.4 to some particularly interesting statistics, including variogram estimators. It follows a simulation study in section 2.5 which compares the performance of the AR sieve bootstrap to normal approximations and the block bootstrap. Section 2.6 contains the proofs of the two central theorems, Baxter's inequality and the result about bootstrap validity, while all other proofs of auxiliary results are deferred to section 2.7.

2.1 Preliminaries

Consider a stationary real-valued spatial process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ with mean zero and finite second moments. In the following we will switch between the two equivalent notations $X_{\underline{t}} = X_{t_1,t_2}$. While the vector index notation $X_{\underline{t}}$ allows for a more compact presentation of the results, the notation X_{t_1,t_2} is sometimes necessary if we want to describe operations on the components of the index vector. For convenience reasons, we will also sometimes use a mixed notation, e.g. in expressions such as $\sum_{t_1\in A}\sum_{t_2\in B} X_{\underline{t}}$. For an introduction to the concept of stationarity for spatial processes, refer to section 1.1.

The autocovariance function of $(X_{\underline{t}})$ at lag $\underline{h} = (h_1, h_2)^T$ is denoted by $\gamma(\underline{h}) = E(X_{\underline{t}+\underline{h}}X_{\underline{t}})$. We assume to have a square-shaped data sample $\{X_{\underline{t}} : 1 \le t_1, t_2 \le n\}$ consisting of n^2 observations at hand. Define $\Pi := \{\underline{t} \in \mathbb{Z}^2 :\ 1 \le t_1, t_2 \le n\}$ and $\Pi_{\underline{h}} := \{\underline{t} \in \mathbb{Z}^2 :\ 1 \le t_1, t_2, t_1 + h_1, t_2 + h_2 \le n\}$; i.e. $\Pi_{\underline{h}}$ describes the set of vectors $\underline{t} \in \mathbb{Z}^2$ such that both $\underline{t}$ and $\underline{t} + \underline{h}$ are elements of Π. The empirical autocovariance function can then be stated as

$$\widehat{\gamma}(\underline{h}) := \frac{1}{|\Pi_{\underline{h}}|}\sum_{\underline{t}\in\Pi_{\underline{h}}}(X_{\underline{t}+\underline{h}} - \overline{X})(X_{\underline{t}} - \overline{X}) \tag{2.1}$$

where $\overline{X} = n^{-2}\sum_{\underline{t}\in\Pi} X_{\underline{t}}$ denotes the sample mean.

We now turn to the algorithm of the autoregressive sieve bootstrap for random fields; for an overview of this concept for time series, see section 1.3. Our proposal depends on fitting an autoregressive model of finite order $p \in \mathbb{N}$ to the data. Since it is not obvious how such an AR fit would look like in the spatial setting, we first define the following set of vectors in $\mathbb{Z}^2$ which characterises the collection of sites for the p-th order AR fit:

$$\Theta(p) := \{\underline{k} \in \mathbb{Z}^2 :\ (1 \le k_1 \le p \text{ and } k_2 = 0) \text{ or } (-p \le k_1 \le p \text{ and } 1 \le k_2 \le p)\}. \tag{2.2}$$

An autoregressive model with sites given by $\Theta(p)$ could be stated as

$$X_{\underline{t}} = \sum_{\underline{k}\in\Theta(p)} a_{\underline{k}}\, X_{\underline{t}-\underline{k}} + e_{\underline{t}} \tag{2.3}$$

for some white noise $(e_{\underline{t}})$. Figure 2.1 illustrates the shape of these types of AR models with an example of order $p = 3$; the index vectors $\underline{t} - \underline{k}$ from (2.3) are marked by the

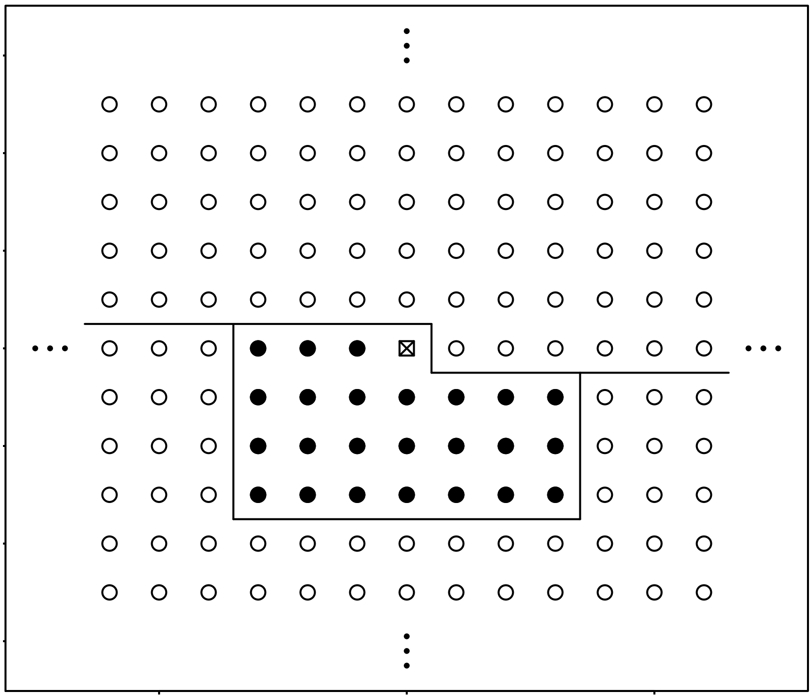

Figure 2.1: Illustration of the shape of an AR(3)-model with respect to $\Theta(3)$, cf. (2.3); locations of sites $\underline{t} - \underline{k}$ marked by the black dots.

black dots while $\underline{t}$ can be found at the center. The AR model from (2.3) is one-sided in the sense of so-called lexicographical ordering of the plane $\mathbb{Z}^2$; we will discuss this property extensively further along the line in this section, but first formulate the AR sieve bootstrap algorithm.

Let $T_n = T_n(\{X_{\underline{t}} : \underline{t} \in \Pi\})$ be an estimator for some unknown parameter θ of the process, based on the given data sample. For an appropriately increasing sequence of real numbers $(c_n)_{n\in\mathbb{N}}$, we assume that the distributions $\mathcal{L}_n = \mathcal{L}(c_n(T_n - \theta))$ converge to a non-degenerated limiting distribution as $n \to \infty$. Our goal is to estimate the distribution $\mathcal{L}_n$ for some finite number $n \in \mathbb{N}$. We propose the following procedure:

The autoregressive sieve bootstrap algorithm for random fields:

(1) Select an order $p = p(n) \in \mathbb{N}$, $p \ll n$ and fit a p-th order autoregressive model of shape (2.3) to the given observations, for example by Yule-Walker estimation. Denote the estimated coefficients by $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$.

(2) Let $\Pi(n,p) := \{(t_1,t_2) \in \mathbb{Z}^2 : p+1 \le t_1 \le n-p,\ p+1 \le t_2 \le n\}$, i.e. $\Pi(n,p)$ is the set of all vectors $\underline{t} \in \Pi$ such that $(\underline{t}-\underline{k}) \in \Pi$ for all $\underline{k} \in \Theta(p)$. Denote the residuals of the autoregressive fit by $\varepsilon'_{\underline{t}}(p) = X_{\underline{t}} - \sum_{\underline{k}\in\Theta(p)} \widehat{a}_{\underline{k}}(p)\, X_{\underline{t}-\underline{k}}$ for all $\underline{t} \in \Pi(n,p)$, and let $\widehat{F}_n$ be the empirical distribution function of the centered residuals $\widehat{\varepsilon}_{\underline{t}}(p) = \varepsilon'_{\underline{t}}(p) - \overline{\varepsilon}$, where $\overline{\varepsilon} = (n-2p)^{-1}(n-p)^{-1}\sum_{\underline{t}\in\Pi(n,p)} \varepsilon'_{\underline{t}}(p)$. Generate independent random variables $\varepsilon^*_{\underline{j}}$ having identical distribution $\widehat{F}_n$, for example by drawing with replacement from the set of centered residuals. Use these resampled residuals and the parameter estimators to calculate a bootstrap sample $\{X^*_{\underline{t}} : \underline{t} \in \Pi\}$ according to the generating equation

$$X^*_{\underline{t}} = \sum_{\underline{k}\in\Theta(p)} \widehat{a}_{\underline{k}}(p)\, X^*_{\underline{t}-\underline{k}} + \varepsilon^*_{\underline{t}}. \tag{2.4}$$

(3) Let $T^*_{n,(1)} := T_n(\{X^*_{\underline{t}} : \underline{t} \in \Pi\})$ be the same estimator as T_n based on the pseudo sample $\{X^*_{\underline{t}} : \underline{t} \in \Pi\}$ and θ^* the analogue of θ associated with the bootstrap process $(X^*_{\underline{t}})$.

(4) Repeat steps (1)–(3) M times, where M is sufficiently large, in order to obtain independent realisations $T^*_{n,(1)}, \ldots, T^*_{n,(M)}$ of the plug-in estimator.

(5) The estimator for $\mathcal{L}_n$ is then given by the empirical distribution of $\mathcal{L}^*_n = \mathcal{L}^*(c_n(T^*_n - \theta^*))$, based on the observations $T^*_{n,(1)}, \ldots, T^*_{n,(M)}$.

Here, $\mathcal{L}^*$ and E^* denote probability law and expectation, conditional on the given data sample.

In the following, we will investigate under which conditions the underlying process $(X_{\underline{t}})$ possesses one-sided autoregressive representations, since this property is crucial for showing asymptotic validity of the AR sieve bootstrap. For the remainder of this chapter we will be working with spatial processes fulfilling the following assumptions. We use the notation $|\underline{k}|_\infty := \max\{|k_1|, |k_2|\}$ for the maximum vector norm of each $\underline{k} \in \mathbb{Z}^2$. For any arbitrary subset A of some vector space over $\mathbb{R}$ or $\mathbb{C}$, $\overline{sp}(A)$ denotes the closed span of all vectors $a \in A$.

Assumption 1. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a strictly stationary real-valued basic process, i.e. $X_{\underline{t}} \notin \overline{sp}\{X_{\underline{s}},\ \underline{s} \neq \underline{t}\}$, with mean zero and finite second moments. The autocovariance function $\gamma(\cdot)$ of $(X_{\underline{t}})$ fulfils $\sum_{\underline{k}\in\mathbb{Z}^2}(1+|\underline{k}|_\infty)^r\,|\gamma(\underline{k})| < \infty$ for some $r \in \mathbb{N}_0$ to be specified in the respective results later on. The spectral density of $(X_{\underline{t}})$,*

$$f(\underline{\lambda}) = \frac{1}{4\pi^2}\sum_{\underline{k}\in\mathbb{Z}^2} \gamma(\underline{k})\, e^{-i\langle \underline{k},\underline{\lambda}\rangle}, \quad \underline{\lambda} \in (-\pi,\pi]^2,$$

fulfils the so-called boundedness condition: There exists a constant $c > 0$ *such that* $f(\underline{\lambda}) \geq c$ *uniformly for all frequencies* $\underline{\lambda} \in (-\pi, \pi]^2$.

Note that this assumption merely requires the spectral density to be positive and smooth, because the weighted summability condition on the autocovariances just implies that certain partial derivatives of f exist. For $u, v \in \mathbb{N}$ with $u + v \leq r$ we get from differentiating the Fourier series of f:

$$\frac{\partial^{u+v} f}{\partial \lambda_1^u \, \partial \lambda_2^v}(\underline{\lambda}) = \frac{1}{4\pi^2} \sum_{\underline{k} \in \mathbb{Z}^2} (-ik_1)^u (-ik_2)^v \, \gamma(\underline{k}) \, e^{-i\langle \underline{k}, \underline{\lambda} \rangle}.$$

The derivative of the Fourier series of f on the right-hand side of the latter equation is absolutely summable because $|(-ik_1)^u(-ik_2)^v| \leq (1 + |\underline{k}|_\infty)^r$ and because of Assumption 1. Therefore, the derivative of f itself, given by the left-hand side, exists and is equal to the derivative of the Fourier series.

We will now establish the aforementioned one-sided autoregressive and moving average representations for all processes that fulfil Assumption 1. Here, *one-sided* refers to the lexicographical ordering of the plane $\mathbb{Z}^2$, cf. Guyon (1995). Defining

$$\Theta := \{(k_1, k_2) \in \mathbb{Z}^2 : \ (k_1 \geq 1 \text{ and } k_2 = 0) \text{ or } (k_1 \text{ arbitrary and } k_2 \geq 1)\}$$

one can observe that $\mathbb{Z}^2$ can be partitioned as $\{\underline{0}\} \cup \Theta \cup (-\Theta)$. Θ is commonly referred to as the *upper half-plane* with respect to the origin while $-\Theta$ is the *lower half-plane*, cf. Helson and Lowdenslager (1958). An illustration is given by Figure 2.2; the upper half-plane Θ is given by the white dots, the lower half-plane by the black dots. Obviously, it holds $\Theta(p) \to \Theta$, as $p \to \infty$.
We now get the following result on one-sided representations for spatial processes:

Lemma 2.1. *Let* $(X_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ *be a spatial process that fulfils Assumption 1 with some* $r \geq 1$. *Then there exist uniquely determined autoregressive (AR) coefficients* $(a_{\underline{k}})_{\underline{k} \in \Theta}$, *uniquely determined moving average (MA) coefficients* $(b_{\underline{k}})_{\underline{k} \in \Theta}$ *and a uniquely determined uncorrelated white noise process* $(\varepsilon_{\underline{t}})$, $\underline{t} \in \mathbb{Z}^2$, *such that* $(X_{\underline{t}})$ *possesses the one-sided AR and MA representations*

$$X_{\underline{t}} = \sum_{\underline{k} \in \Theta} a_{\underline{k}} \, X_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}}, \quad X_{\underline{t}} = \sum_{\underline{k} \in \Theta} b_{\underline{k}} \, \varepsilon_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}}, \tag{2.5}$$

respectively, and $\sum_{\underline{k} \in \Theta} a_{\underline{k}} \, X_{\underline{t}-\underline{k}}$ *represents the* L^2*-projection of* $X_{\underline{t}}$ *onto* $\overline{sp}\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta\}$. *The white noise process* $(\varepsilon_{\underline{t}})$ *is called the innovation process of* $(X_{\underline{t}})$. *The* *coefficients in* (2.5) *fulfil the summability conditions*

$$\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^{r-1}\,|a_{\underline{k}}| < \infty, \quad \sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^{r-1}\,|b_{\underline{k}}| < \infty. \tag{2.6}$$

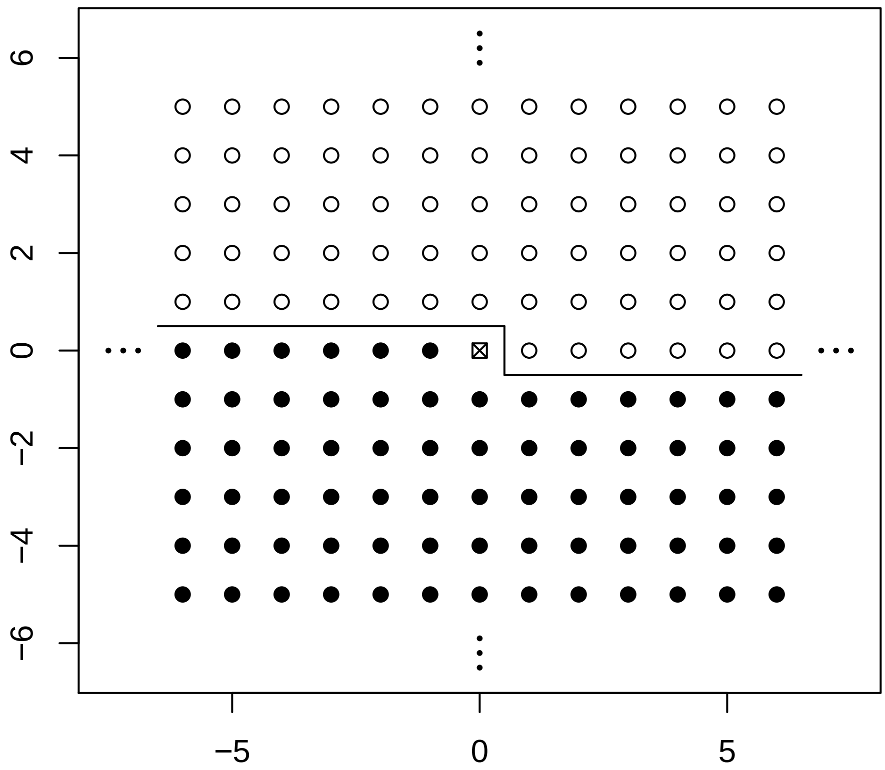

Figure 2.2: Illustration of the upper (white dots) and lower (black dots) half-plane of $\mathbb{Z}^2$.

It should be noted that the existence of representations (2.5) has already been proven by Whittle (1954). However, we are especially interested in the summability conditions (2.6), which are not available in the literature. Hence, we derive these conditions in the proof of Lemma 2.1, which can be found in section 2.7.

Remark 2.2. At this point we should clarify a common misunderstanding in the discussion of spatial and time series autoregressions: For time series, the 'past' and the 'future' of a time value $t \in \mathbb{Z}$ are naturally defined, and it is generally accepted that random variables X_t are influenced by its past values $X_{t-1}, X_{t-2}, \ldots$ as in (1.2). Since this is not the case for spatial processes, it is often criticized that one has to choose a concept of 'past' values, i.e. choose a direction from which the random variable $X_{\underline{t}}$ is influenced, such as the lower half-plane illustrated by Figure 2.2. This choice is of course arbitrary, which is why one might come to the conclusion that the whole concept of one-sided autoregressions implies a very specific model assumption which is not fulfilled for real-world data. However, the opposite is true: The AR sieve bootstrap, as an example, only uses the one-sided autoregressions as a vehicle in the proof of bootstrap validity. Under the mild conditions from Assumption 1, which only depend on the spectral density and which do not include any choice of direction whatsoever, the process $(X_{\underline{t}})$ possesses autoregressive representations with respect to *each half-plane* of $\mathbb{Z}^2$ that might be chosen. Therefore, the whole procedure is by no means arbitrary; and the concept of approximating a particular one-sided autogression does not constrain the class of processes any further than demanding the spectral density to be positive and smooth. □

In order to prove the summability conditions from Lemma 2.1 we need the following auxiliary result. The AR and MA coefficients are strongly connected to the so-called *cepstral coefficients* of the process, that is the Fourier coefficients of the logarithm of the spectral density. The following lemma provides a result that carries over the summability condition from the Fourier coefficients of a function f to the Fourier coefficients of its logarithm. The result holds not only for spectral densities but for arbitrary integrable functions, and seems not to be available in the literature so far, at least not in this explicit form.

Lemma 2.3. *Denote for every integrable function $f : (-\pi, \pi]^2 \to \mathbb{R}$ its Fourier coefficients by $\widetilde{f}_{\underline{k}} = (1/4\pi^2) \int_{(-\pi,\pi]^2} f(\underline{\lambda})\, e^{-i\langle \underline{k},\underline{\lambda}\rangle}\, d\underline{\lambda}$ and its formal Fourier series by $\sum_{\underline{k}\in\mathbb{Z}^2} \widetilde{f}_{\underline{k}}\, e^{i\langle \underline{k},\underline{\lambda}\rangle}$. We define the following classes of functions:*

$$\begin{aligned} C_r &:= \Big\{ f : (-\pi,\pi]^2 \to \mathbb{R},\ \|f\|_r := \sum_{\underline{k}\in\mathbb{Z}^2} (1+|\underline{k}|_\infty)^r\, |\widetilde{f}_{\underline{k}}| < \infty \Big\}, \\ D_{r_1,r_2} &:= \Big\{ f : (-\pi,\pi]^2 \to \mathbb{R},\ \|f\|_{r_1,r_2} := \sum_{\underline{k}\in\mathbb{Z}^2} (1+|k_1|)^{r_1}(1+|k_2|)^{r_2}\, |\widetilde{f}_{\underline{k}}| < \infty \Big\}. \end{aligned}$$

Assume that $f(\underline{\lambda}) \geq c > 0$ for all $\underline{\lambda} \in (-\pi,\pi]^2$. Then it holds:

(i) If $f \in C_r$ for some $r \geq 2$, it follows $\log f \in C_{r-1}$.

(ii) If $f \in D_{r_1,r_2}$ for some $r_1, r_2 \geq 1$, it follows $\log f \in D_{r_1,r_2}$.

Remark 2.4. In Assumption 1 and Lemma 2.3 (i) we use the weight function $\nu(\underline{k}) = (1+|\underline{k}|_\infty)^r$. This is due to the fact that we will later establish a weighted version of a Baxter-inequality for spatial processes, cf. Theorem 2.7. The proof of this Baxter-inequality requires the weights to be strictly non-decreasing in $|\underline{k}|_\infty$, i.e. $\nu(\underline{k}) \geq \nu(\underline{j})$ whenever $|\underline{k}|_\infty \geq |\underline{j}|_\infty$, which is why we chose this particular weight function. Other weights one might think of, like replacing the $|\cdot|_\infty$-norm in $\nu(\underline{k})$ by the Euclidean norm, the 1-norm or letting $\widetilde{\nu}(\underline{k}) = (1+|k_1|)^{r_1}(1+|k_2|)^{r_2}$, do not fulfil the property of being strictly non-decreasing in $|\underline{k}|_\infty$ and are, therefore, not suitable in order to establish a weighted Baxter-inequality. However, for Assumption 1 to be fulfilled, it suffices to check whether $\sum_{\underline{k}\in\mathbb{Z}^2}(1+|\underline{k}|)^r\,|\gamma(\underline{k})| < \infty$ for any vector norm $|\underline{k}|$, since all vector norms are equivalent. □

Remark 2.5. Classes of functions with weighted absolutely summable Fourier coefficients, such as C_r and D_{r_1,r_2} from Lemma 2.3, are commonly referred to as *Beurling algebras*; C_r represents the special case for the weight function $\nu(\underline{k}) = (1+|\underline{k}|_\infty)^r$. Remark 2.4 explains why we are looking at these particular weights, although we get the somehow unsatisfactory result that $f \in C_r$ does *not imply* $\log f \in C_r$, but instead $\log f \in C_{r-1}$. While we will only work with assertion (i) from Lemma 2.3 for the remainder of this thesis, it is still worthwile to consider the class D_{r_1,r_2} from (ii). Here, we get with analogous arguments as in (i) that $f \in D_{r_1,r_2}$ implies $\log f \in D_{r_1,r_2}$, i.e. the Fourier coefficients of $\log f$ fulfil the same summability condition as the ones of f. This result is strongly connected to the well-known Wiener-Lévy-Theorem (cf. Zygmund (2002), Chapter VI, Theorem 5.2); and, for the special case of $\phi(f) = \log f$, our result even represents a slight generalisation of the latter, with respect to functions in several variables. We will shed some light on this situation:

Originally, Norbert Wiener proved for functions in one variable that if $f \neq 0$ has absolutely summable Fourier coefficients, then the same holds true for $1/f$. This assertion, also known as Wiener's lemma, can be transferred to functions in several variables; and, moreover, weighted summability versions in the spirit of Lemma 2.3 are available, cf. Theorem 6.2 in Gröchenig (2007). For functions in one variable, Paul Lévy generalised Wiener's result, concluding that if f has absolutely summable Fourier coefficients, the same holds true for $\phi(f)$, where ϕ is a smooth functional. This assertion became known as the Wiener-Lévy-Theorem. In contrast to what happens for $\phi(f) = 1/f$, weighted versions in several variables are much harder to come by for general functions ϕ. Typically, one only gets that $\phi(f)$ is the element of a Beurling algebra with weights increasing at a slower rate than the ones of f, cf. Bhatt and Dedania (2003).

Our proof of Lemma 2.3 (ii) shows that a generalisation to functions in several variables for the special case of $\phi(f) = \log f$ is possible. However, the proof relies heavily on the structure of the logarithmic function and cannot be generalised to other functions. □

2.2 Convergence of finite-order model fits

In this section we will establish results that ensure convergence of the estimated parameters $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ from step (1) of the AR sieve bootstrap procedure, cf. section 2.1, towards the autoregressive coefficients $\{a_{\underline{k}} : \underline{k} \in \Theta\}$ of the underlying process given by Lemma 2.1. We will split up the results in two subsections: The first one will be concerned with convergence of the finite predictor coefficients of the process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ towards $\{a_{\underline{k}} : \underline{k} \in \Theta\}$. The finite predictors are the L^2-projection coefficients of random variable $X_{\underline{t}}$ to the finite-dimensional space $sp\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta(p)\}$. Here, if A is an arbitrary subset of some vector space over $\mathbb{R}$ or $\mathbb{C}$, $sp(A)$ denotes the span of all vectors $a \in A$. In this context we will introduce a Baxter-inequality for random fields. Section 2.2.2 deals with conditions which ensure that the difference between the estimators $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ and the finite predictor coefficients vanishes asymptotically in probability. The results from both subsections combined then yield the desired convergence of the finite-order AR model fits.

2.2.1 Convergence of finite predictor coefficients

The finite predictor coefficients with respect to the set $\Theta(p)$ are the coefficients of the L^2-projection of $X_{\underline{t}}$ onto $sp\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta(p)\}$, and will be denoted by $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$. They can be obtained from solving the minimization problem

$$\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\} := \underset{\left\{c_{\underline{k}}(p):\, \underline{k}\in\Theta(p)\right\}}{\arg\min}\; E\left(X_{\underline{t}} - \sum_{\underline{k}\in\Theta(p)} c_{\underline{k}}(p)\, X_{\underline{t}-\underline{k}}\right)^2 . \tag{2.7}$$

Solving (2.7) leads to the well-known Yule-Walker equations. We now want to introduce the notation which allows us to write the Yule-Walker equations in a convenient form: The number of elements in $\Theta(p)$ is $\bar{p} := 2p(p+1)$. Let $\underline{k}_1, \ldots, \underline{k}_{\bar{p}}$ be an arbitrary enumeration of the vectors $\underline{k} \in \Theta(p)$. Define $\underline{a}(p) := (a_{\underline{k}_1}(p), \ldots, a_{\underline{k}_{\bar{p}}}(p))^T \in \mathbb{R}^{\bar{p}}$ and $\underline{Y}_{\underline{t}} := (X_{\underline{t}-\underline{k}_1}, \ldots, X_{\underline{t}-\underline{k}_{\bar{p}}})^T$. Note that the indices $\underline{k}_j$ appear in the same order in both vectors. Due to the projection property it is easy to see that any solution of (2.7) fulfils

$$E\left((X_{\underline{t}} - \underline{a}(p)^T \underline{Y}_{\underline{t}}) \cdot \underline{Y}_{\underline{t}}^T\, \underline{e}_j\right) = 0, \quad j = 1, \ldots, \bar{p}, \tag{2.8}$$

where $\underline{e}_j$ denotes the j-th unit vector. Using the notation $\Gamma(p) := E(\underline{Y}_{\underline{t}}\underline{Y}_{\underline{t}}^T)$ and $\underline{\gamma}(p) := E(X_{\underline{t}}\underline{Y}_{\underline{t}})$, system (2.8) is equivalent to

$$\Gamma(p)\,\underline{a}(p) = \begin{pmatrix} \gamma(\underline{k}_1 - \underline{k}_1) & \cdots & \gamma(\underline{k}_1 - \underline{k}_{\bar{p}}) \\ \vdots & \ddots & \vdots \\ \gamma(\underline{k}_{\bar{p}} - \underline{k}_1) & \cdots & \gamma(\underline{k}_{\bar{p}} - \underline{k}_{\bar{p}}) \end{pmatrix} \cdot \begin{pmatrix} a_{\underline{k}_1}(p) \\ \vdots \\ a_{\underline{k}_{\bar{p}}}(p) \end{pmatrix} = \begin{pmatrix} \gamma(\underline{k}_1) \\ \vdots \\ \gamma(\underline{k}_{\bar{p}}) \end{pmatrix} = \underline{\gamma}(p). \quad (2.9)$$

System (2.9) is called the Yule-Walker equations. Note that the matrix $\Gamma(p)$ is symmetric, regardless of the order of indices in the vectors $\underline{Y}_{\underline{t}}$ and $\underline{a}(p)$. The following result ensures the existence of a unique solution of (2.9). Moreover, we establish a uniform bound for the spectral norms of the inverse matrices $\Gamma(p)^{-1}$, which will turn out to be crucial for proving the Baxter-inequality. The spectral norm of a real-valued quadratic matrix A is defined as the square root of the largest eigenvalue of A^TA, denoted by $\|A\|_{spec} = \sqrt{\sigma_{max}(A^TA)}$. For symmetric positive definite matrices this formula can be simplified to $\|A\|_{spec} = \sqrt{\sigma_{max}(A^TA)} = \sigma_{max}(A)$.

Lemma 2.6. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process that fulfils Assumption 1. Then the matrix $\Gamma(p)$ from the Yule-Walker equations (2.9) is invertible for all $p \in \mathbb{N}$. Furthermore, it holds $\|\Gamma(p)^{-1}\|_{spec} \leq (4\pi^2 c)^{-1}$ for all $p \in \mathbb{N}$, where c is the lower bound of the spectral density from Assumption 1, and $\|\cdot\|_{spec}$ denotes the spectral norm.*

The previous lemma justifies calling the unique solution $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ of (2.9) *the* finite predictor coefficients of the process for order p. As already mentioned, it is of critical importance for our sieve bootstrap scheme that the $a_{\underline{k}}(p)$ converge towards the autoregressive coefficients $\{a_{\underline{k}} : \underline{k} \in \Theta\}$ of the underlying process from (2.5), as p tends to infinity. In particular, we have to ensure that this convergence is fast enough. Therefore, we introduce the following version of Baxter's inequality for random fields:

Theorem 2.7. (Baxter's Inequality) *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process that fulfils Assumption 1 with some $r \geq 2$ and $c > 0$. Let $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ be its finite predictor coefficients as defined above, and $\{a_{\underline{k}} : \underline{k} \in \Theta\}$ be its autoregressive coefficients given by* (2.5)*. Denote by $K := \sum_{\underline{k}\in\mathbb{Z}^2} |\gamma(\underline{k})|$. Then it holds for all $s \in \mathbb{N}_0$ with $s+1 < r$ and for all $p \in \mathbb{N}$:*

$$\sum_{\underline{k}\in\Theta(p)} (1 + |\underline{k}|_\infty)^s \, |a_{\underline{k}}(p) - a_{\underline{k}}| \leq \frac{K}{2\sqrt{2}\,\pi^2 c} \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1 + |\underline{k}|_\infty)^{s+1} \, |a_{\underline{k}}|.$$

Due to Lemma 2.1 the right-hand side converges to zero as $p \to \infty$.

The established convergence of the autoregressive coefficients in Baxter's inequality is closely related to a similar convergence of moving average parameters, which shall be derived in the next step. To do this, we take a look at so-called *z-transforms*, also called *transfer functions*, cf. Brockwell and Davis (1991), section 4.4. Based on the AR and MA representations from (2.5) with the coefficients $(a_{\underline{k}})$ and $(b_{\underline{k}})$, we define the z-transforms

$$A(\underline{z}) = 1 - \sum_{\underline{k}\in\Theta} a_{\underline{k}}\, z_1^{k_1} z_2^{k_2}, \quad B(\underline{z}) = 1 + \sum_{\underline{k}\in\Theta} b_{\underline{k}}\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S, \tag{2.10}$$

where

$$S := \{\underline{z} \in \mathbb{C}^2 : \ |z_1| = 1,\ |z_2| \leq 1\}.$$

The series $A(\underline{z})$ and $B(\underline{z})$ converge absolutely on its domain S because of Lemma 2.1. It is worth noting that we have to make the distinction between z_1 and z_2 in S. Since z_2 shows up exclusively with exponents $k_2 \geq 0$ in (2.10), as can be seen from the definition of Θ in section 2.1, we have $|z_2|^{k_2} \leq 1$ for the entire closed disk $|z_2| \leq 1$, while z_1 shows up with both positive and negative exponents k_1. Hence we get $|z_1|^{k_1} \leq 1$, and thus absolute convergence of the series $A(\underline{z})$ and $B(\underline{z})$, only for the circle $|z_1| = 1$.

In analogy to the definition of $A(\underline{z})$, we now define the z-transform of the finite predictor coefficients $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ by

$$A_p(\underline{z}) = 1 - \sum_{\underline{k}\in\Theta(p)} a_{\underline{k}}(p)\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S_p, \tag{2.11}$$

where

$$S_p := \left\{\underline{z} \in \mathbb{C}^2 : \ \frac{p}{p+1} \leq |z_1| \leq \frac{p+1}{p},\ 0 \leq |z_2| \leq \frac{p+1}{p}\right\}.$$

Note that $A_p(\underline{z})$ is defined on an extended domain compared to $A(\underline{z})$, but for $p \to \infty$ the domains S_p converge to S.

From the proof of Lemma 2.1 we already have $B(\underline{z}) = 1/A(\underline{z})$ for all $\underline{z} \in S$. In particular, both $A(\underline{z})$ and $B(\underline{z})$ are non-zero on their domain S. The next lemma shows that, for p large enough, the inverse of $A_p(\underline{z})$ has a z-transform similar to the one of $B(\underline{z})$.

Lemma 2.8. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process that fulfils the conditions of Theorem 2.7 with some $r \geq 2$. Then there exists $\delta > 0$ such that it holds $|A_p(\underline{z})| \geq \delta$ uniformly for all $\underline{z} \in S_p$ and all p large enough. For those p, $B_p(\underline{z}) := 1/A_p(\underline{z})$ can be expressed as a convergent series of the form*

$$B_p(\underline{z}) = 1 + \sum_{\underline{k}\in\Theta} b_{\underline{k}}(p)\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S_p, \tag{2.12}$$

for suitable coefficients $\{b_{\underline{k}}(p) : \underline{k} \in \Theta\}$.

We conclude this section with a result which transfers the convergence of the autoregressive parameters from Baxter's inequality to the moving average parameters $\{b_{\underline{k}}(p) : \underline{k} \in \Theta\}$ and $\{b_{\underline{k}} : \underline{k} \in \Theta\}$:

Lemma 2.9. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process that fulfils the conditions of Theorem 2.7 with some $r \geq 2$. For all p large enough such that $A_p(\underline{z}) \neq 0$ for all $\underline{z} \in S_p$, let $\{b_{\underline{k}}(p) : \underline{k} \in \Theta\}$ be the coefficients as defined in (2.12) and let $(a_{\underline{k}})_{\underline{k}\in\Theta}$ and $(b_{\underline{k}})_{\underline{k}\in\Theta}$ be the AR and MA coefficients of $(X_{\underline{t}})$ given by (2.5). Then there exists a constant $C < \infty$ such that it holds for all p large enough, and for all $s \in \mathbb{N}_0$ with $s + 1 < r$:*

$$\sum_{\underline{k}\in\Theta} (1 + |\underline{k}|_\infty)^s\, |b_{\underline{k}}(p) - b_{\underline{k}}| \leq C \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1 + |\underline{k}|_\infty)^{s+1}\, |a_{\underline{k}}|.$$

Due to Lemma 2.1, the right-hand side converges to zero as $p \to \infty$.

The proofs for all lemmas in this section can be found in section 2.7, except for Theorem 2.7, which can be found in section 2.6.

2.2.2 Conditions on the fitted-model order $p(n)$ and convergence of estimated coefficients

It is important for the validity of the AR sieve bootstrap scheme that the parameter estimators $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ used in step 1 of the procedure converge towards the finite predictor coefficients $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ at a sufficient rate. At this point one has to keep in mind that the order p of the autoregressive fits actually depends on the sample size n, which is suppressed in the notation for most parts of this thesis due to convenience reasons. In order to use the results from the previous section, we need $p = p(n) \to \infty$ as $n \to \infty$. This implies that the dimension of the Yule-Walker matrices $\Gamma(p)$ given by (2.9) also increases for $n \to \infty$.

Probably the most popular form of fitting an AR model as in step (1) of the sieve bootstrap procedure, is Yule-Walker estimation: One replaces the autocovariances in $\Gamma(p)$ by its empirical versions, cf. (2.1), and solves the linear system. Informally speaking, we then have to make sure that $p(n)$ increases slowly enough such that for n large enough all autocovariances showing up in $\Gamma(p)$ can be estimated sufficiently well, in order to obtain a small difference between $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ and $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$.

The following assumption formalizes this condition. Essentially it contains two assertions: Firstly, the underlying process allows for consistent estimation of the finite predictor coefficients $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$. Secondly, by restricting the rate of increase of $p = p(n)$, we can achieve sufficiently fast uniform convergence of the estimators $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$.

Assumption 2. *For $p = p(n)$, with $p(n) \to \infty$ as $n \to \infty$, assume for the following sequence in n:*

$$p^4 \cdot \sum_{\underline{k} \in \Theta(p)} |\widehat{a}_{\underline{k}}(p) - a_{\underline{k}}(p)| = \mathcal{O}_P(1).$$

In the remainder of this section we will investigate whether the fitted AR models can also be represented as moving averages of possibly infinite order, which will be crucial for asymptotic inference later on. Based on the parameter estimators $\widehat{a}_{\underline{k}}(p)$ we can define the z-transform $\widehat{A}_p(\underline{z})$ analogously to $A_p(\underline{z})$ in (2.11) as

$$\widehat{A}_p(\underline{z}) = 1 - \sum_{\underline{k} \in \Theta(p)} \widehat{a}_{\underline{k}}(p)\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S_p.$$

The following calculations will make sure that $\widehat{A}_p(\underline{z})$ is bounded away from zero for n large enough. Assumption 2 implies

$$\begin{aligned} \sup_{\underline{z} \in S_p} \left|\widehat{A}_p(\underline{z}) - A_p(\underline{z})\right| &\leq \sum_{\underline{k} \in \Theta(p)} \left|\widehat{a}_{\underline{k}}(p) - a_{\underline{k}}(p)\right| \left(\frac{p+1}{p}\right)^{|k_1|+k_2} \\ &\leq \left(\frac{p+1}{p}\right)^{2p} \sum_{\underline{k} \in \Theta(p)} |\widehat{a}_{\underline{k}}(p) - a_{\underline{k}}(p)| \\ &= \frac{1}{p^4}\, \mathcal{O}_P(1) = o_P(1), \end{aligned} \tag{2.13}$$

because $((p+1)/p)^{2p}$ is a bounded sequence (convergent with limit e^2), and because the definition of S_p yields

$$\begin{aligned} |z_1|^{k_1} &\le \left\{ \begin{matrix} \left(\frac{p+1}{p}\right)^{k_1}, & \text{for } k_1 \ge 0 \\ \left(\frac{p}{p+1}\right)^{k_1}, & \text{for } k_1 < 0 \end{matrix} \right\} = \left(\frac{p+1}{p}\right)^{|k_1|}, \\ |z_2|^{k_2} &\le \left(\frac{p+1}{p}\right)^{k_2}, \end{aligned}$$

for all $\underline{z} \in S_p$. Assumption 2 ensures $p \to \infty$, as $n \to \infty$, which implies that $A_p(\underline{z})$ is bounded away from zero for all n large enough, cf. Lemma 2.8. It follows from (2.13) that $\widehat{A}_p(\underline{z})$ is uniformly bounded away from zero in probability for all $\underline{z} \in S_p$ and for all n large enough. For all those n large enough, the inverse of $\widehat{A}_p(\underline{z})$ possesses the expansion

$$\widehat{B}_p(\underline{z}) = \frac{1}{\widehat{A}_p(\underline{z})} = 1 + \sum_{\underline{k}\in\Theta} \widehat{b}_{\underline{k}}(p)\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S_p, \tag{2.14}$$

in probability, following the same arguments as for (2.12). Hence, the bootstrap process given by (2.4), which can be described by the transfer function $\widehat{A}_p(\underline{z})$, has the moving average representation

$$X^*_{\underline{t}} = \sum_{\underline{k}\in\Theta} \widehat{b}_{\underline{k}}(p)\, \varepsilon^*_{\underline{t}-\underline{k}} + \varepsilon^*_{\underline{t}} \tag{2.15}$$

for all n large enough, in probability. The convergence of the parameter estimators $\widehat{a}_{\underline{k}}(p)$ towards $a_{\underline{k}}(p)$ in Assumption 2 carries over to the corresponding moving average parameters, as shows the following lemma.

Lemma 2.10. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process that fulfils the conditions of Theorem 2.7 and Assumption 2. Then, for all n large enough (and thus p large enough) such that $A_p(\underline{z})$ and $\widehat{A}_p(\underline{z})$ are bounded away from zero (the latter in probability), it holds uniformly for all $\underline{k} \in \Theta$ and for some $C < \infty$:*

$$\left|\widehat{b}_{\underline{k}}(p) - b_{\underline{k}}(p)\right| \le C \cdot \left(1 + \frac{1}{p}\right)^{-|k_1|-k_2} \frac{1}{p^4} \quad \textit{in probability.}$$

The proof can be found in section 2.7.

2.3 Asymptotic validity of the bootstrap

In this section we will derive asymptotic validity of the AR sieve bootstrap procedure under appropriate conditions for a class of statistics which will be specified in

Assumption 3. Similar to what happens in the time series case, cf. Kreiss, Paparoditis and Politis (2011), it turns out that the bootstrap procedure asymptotically mimics the behaviour of the so-called companion process, a modification of the underlying process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$. This yields a check criterion which basically says that the bootstrap procedure works asymptotically for a test statistic T_n, whenever the asymptotic distributions of T_n applied to the underlying and the companion process coincide. We will elaborate this, and start with the definition of the companion process:

Based on representation (2.5) for the underlying process, we define the *companion process* of $(X_{\underline{t}})$ as the stationary spatial process $(\widetilde{X}_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$, generated by

$$\widetilde{X}_{\underline{t}} = \sum_{\underline{k}\in\Theta} a_{\underline{k}}\, \widetilde{X}_{\underline{t}-\underline{k}} + \widetilde{\varepsilon}_{\underline{t}}, \tag{2.16}$$

where the coefficients $a_{\underline{k}}$ are exactly the ones from (2.5) and $(\widetilde{\varepsilon}_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ is an i.i.d. white noise process with identical marginal distribution as $(\varepsilon_{\underline{t}})$, i.e. $\mathcal{L}(\widetilde{\varepsilon}_{\underline{t}}) = \mathcal{L}(\varepsilon_{\underline{t}})$. Therefore, the companion process also possesses the moving average representation

$$\widetilde{X}_{\underline{t}} = \sum_{\underline{k}\in\Theta} b_{\underline{k}}\, \widetilde{\varepsilon}_{\underline{t}-\underline{k}} + \widetilde{\varepsilon}_{\underline{t}}, \tag{2.17}$$

with the exact same coefficients $b_{\underline{k}}$ as in (2.5). The only difference between $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$ is the dependence structure of the respective noise processes $(\varepsilon_{\underline{t}})$ and $(\widetilde{\varepsilon}_{\underline{t}})$. While $(\widetilde{\varepsilon}_{\underline{t}})$ is i.i.d., $(\varepsilon_{\underline{t}})$ is strictly stationary but *not necessarily independent*, the random variables $\varepsilon_{\underline{s}}$ and $\varepsilon_{\underline{t}}$ in general are only uncorrelated for $\underline{s} \neq \underline{t}$. Nevertheless, it is easy to see from (2.17) that all second order properties of $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$ are identical, i.e. the two processes possess identical autocovariances and spectral densities.

In our main theorem we will establish bootstrap validity for a class of statistics which will be specified in the following Assumption 3. This class is a natural extension of the so-called *functions of generalized means*, introduced by Künsch (1989), to the case of random fields. These statistics will be based on smooth functions g applied to rectangular-shaped subsamples of the available data sample $\{X_{\underline{t}} : \underline{t} \in \Pi\}$, with $\Pi := \{\underline{t} \in \mathbb{Z}^2 : \ 1 \leq t_1, t_2 \leq n\}$. We first specify the necessary notation: For $1 \leq m_1, m_2 \leq n$ let

$$\begin{aligned} S(m_1, m_2) : &= \left\{\underline{s} = (s_1, s_2)^T \in \mathbb{N}_0^2 : \ 0 \leq s_1 \leq m_1 - 1, \quad 0 \leq s_2 \leq m_2 - 1\right\} \\ &= \{\underline{s}(1), \ldots, \underline{s}(m_1 m_2)\}, \end{aligned}$$

i.e. $\underline{s}(1), \ldots, \underline{s}(m_1 m_2)$ is any fixed enumeration of the $m_1 m_2$ vectors in $S(m_1, m_2)$. We define the $m_1 m_2$-dimensional random vector

$$\mathbf{Y}_{\underline{t}} := (X_{\underline{t}+\underline{s}(1)}, \ldots, X_{\underline{t}+\underline{s}(m_1 m_2)})^T.$$

Observe that for each $\underline{t}$ with $1 \leq t_1 \leq n - m_1 + 1$ and $1 \leq t_2 \leq n - m_2 + 1$, the components of $\mathbf{Y}_{\underline{t}}$ form a rectangular-shaped subsample of dimension $m_1 \times m_2$ of the original data sample. We can now specify the class of statistics we will be investigating.

Assumption 3. *Let $\bar{n}_1 := n - m_1 + 1$, $\bar{n}_2 := n - m_2 + 1$ for some $1 \leq m_1, m_2 \leq n$, and let $m := m_1 m_2$. Define the statistic T_n as*

$$T_n = f\left(\frac{1}{\bar{n}_1 \bar{n}_2} \sum_{t_1=1}^{\bar{n}_1} \sum_{t_2=1}^{\bar{n}_2} g\Big(\mathbf{Y}_{\underline{t}}\Big)\right)$$

where the functions $g : \mathbb{R}^m \to \mathbb{R}^k$ and $f : \mathbb{R}^k \to \mathbb{R}$, with $k \geq 1$, fulfil the following smoothness conditions: f is continuously differentiable in a neighborhood of $\underline{\theta} := E\, g(\mathbf{Y}_{\underline{t}})$ and the gradient of f at $\underline{\theta}$ does not vanish, i.e.

$$\nabla f(\theta) = \Big(\frac{\partial f(\underline{x})}{\partial x_1}, \ldots, \frac{\partial f(\underline{x})}{\partial x_k}\Big)\Big|_{\underline{x}=\underline{\theta}} \neq (0, \ldots, 0).$$

For some $h \geq 1$ all component functions $g_1, \ldots, g_k$ of g are h times continuously differentiable and all h-th-order derivatives satisfy a Lipschitz condition, i.e. for all $i = 1, \ldots, k$ and for all $(h_1, \ldots, h_m) \in \mathbb{N}_0^m$ with $\sum_{u=1}^m h_u = h$ the derivative

$$\frac{\partial^h g_i(\underline{x})}{\partial^{h_1} x_1 \ldots \partial^{h_m} x_m}$$

is Lipschitz.

Remark 2.11. The conditions from the previous assumption should be explained at this point: The class of statistics from Assumption 3 contains, among other things, the sample mean and versions of the sample autocovariance and sample autocorrelation. To obtain the latter two statistics, one typically uses a function g which is not Lipschitz. For example, in the case of sample autocovariances at lag $\underline{h} = (h_1, h_2)^T$, one may choose $m_1 = h_1 + 1$, $m_2 = h_2 + 1$ and $g(x_1, \ldots, x_m) = x_1 x_m$. Then T_n from Assumption 3 translates to taking the empirical mean of observations $X_{\underline{t}+\underline{h}} X_{\underline{t}}$. Now observe that g itself is *not Lipschitz*, but all of its first order partial derivatives are. This is the why we allow for non-Lipschitz functions g in Assumption 3, and merely assume that there exists a number $1 \leq h < \infty$ such that all derivatives of order h (but *not* up to order h) are Lipschitz. $\square$

In order to state the main theorem, we define $\widetilde{T}_n$ and T_n^* as the statistic T_n applied to samples from the companion process $(\widetilde{X}_{\underline{t}})$ and the bootstrap process $(X^*_{\underline{t}})$, respectively, i.e.

$$\widetilde{T}_n := f\left(\frac{1}{\bar{n}_1\bar{n}_2}\sum_{t_1=1}^{\bar{n}_1}\sum_{t_2=1}^{\bar{n}_2} g\Big(\widetilde{\mathbf{Y}}_{\underline{t}}\Big)\right), \quad T_n^* := f\left(\frac{1}{\bar{n}_1\bar{n}_2}\sum_{t_1=1}^{\bar{n}_1}\sum_{t_2=1}^{\bar{n}_2} g\Big(\mathbf{Y}^*_{\underline{t}}\Big)\right)$$

where

$$\widetilde{\mathbf{Y}}_{\underline{t}} := (\widetilde{X}_{\underline{t}+\underline{s}(1)}, \dots, \widetilde{X}_{\underline{t}+\underline{s}(m_1m_2)})^T, \quad \mathbf{Y}^*_{\underline{t}} := (X^*_{\underline{t}+\underline{s}(1)}, \dots, X^*_{\underline{t}+\underline{s}(m_1m_2)})^T.$$

We can prove bootstrap validity under the following assumptions, which ensure convergence of empirical moments and the empirical distribution function to their theoretical counterparts for the innovations:

Assumption 4. *For all continuity points $x \in \mathbb{R}$ of the distribution function F of $\varepsilon_{\underline{0}}$ it holds*

$$F_n(x) \xrightarrow{P} F(x) \quad \text{as } n \to \infty,$$

where $F_n(x)$ is the empirical distribution function

$$F_n(x) = \frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)} \mathbb{1}\{\varepsilon_{\underline{t}} \le x\},$$

and where $\Pi(n,p) := \{(t_1,t_2) \in \mathbb{Z}^2 : p+1 \le t_1 \le n-p,\ p+1 \le t_2 \le n\}$.
Furthermore, it holds $E\Big(\varepsilon_{\underline{t}}^{2(h+2)}\Big) < \infty$, where h is the constant specified in Assumption 3, as well as the following convergence of empirical moments:

$$\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)} (\varepsilon_{\underline{t}})^{2w} \xrightarrow{P} E\Big((\varepsilon_{\underline{0}})^{2w}\Big) \quad \forall\, w \le h+2.$$

Theorem 2.12. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process fulfilling Assumptions 2 – 4, as well as Assumption 1 with $r = 4$.*
Then, for $\widetilde{T}_n$ and T_n^* as defined above, it holds

$$d_K\Big(\mathcal{L}^*\Big(n\,(T_n^* - f(\underline{\theta}^*))\Big), \mathcal{L}\Big(n\,(\widetilde{T}_n - f(\underline{\widetilde{\theta}}))\Big)\Big) = o_P(1)$$

as $n \to \infty$, where $\underline{\theta}^* = E^*\Big(g(\mathbf{Y}^*_{\underline{t}})\Big)$, $\underline{\widetilde{\theta}} = E\Big(g(\widetilde{\mathbf{Y}}_{\underline{t}})\Big)$ and d_K denotes the Kolmogorov distance.

This result shows for all statistics from Assumption 3 that the sieve bootstrap procedure asymptotically approximates the distribution $\widetilde{T}_n$ instead of the one of T_n. Therefore, the bootstrap procedure works asymptotically *if and only if* the limiting distributions of T_n and $\widetilde{T}_n$ coincide. We will give a few examples of the application of this check criterion in the following section.

The proof of Theorem 2.12 can be found in Section 2.6.2 and depends in large parts on some auxiliary results that will be collected in the following lemmas. We will make use of a truncated version $(X^*_{\underline{t},M})$ of the bootstrap process, which is based on the moving average representation of $(X^*_{\underline{t}})$ from (2.15). For arbitrary $M \in \mathbb{N}$ we define

$$X^*_{\underline{t},M} = \sum_{\underline{k}\in\Theta(M)} \widehat{b}_{\underline{k}}(p)\, \varepsilon^*_{\underline{t}-\underline{k}} + \varepsilon^*_{\underline{t}}, \tag{2.18}$$

where the finite collection of sites $\Theta(M)$ is defined in (2.2), whereas the non-truncated version $(X^*_{\underline{t}})$ has the infinite collection of sites Θ. Analogously, a truncated version $(\widetilde{X}_{\underline{t},M})$ of the companion process can be defined by replacing Θ with $\Theta(M)$ in (2.17). As a natural extension of the definition of $\mathbf{Y}^*_{\underline{t}}$ and $\widetilde{\mathbf{Y}}_{\underline{t}}$, we denote by

$$\mathbf{Y}^*_{\underline{t},M} := (X^*_{\underline{t}+\underline{s}(1),M}, \ldots, X^*_{\underline{t}+\underline{s}(m_1m_2),M})^T, \quad \widetilde{\mathbf{Y}}_{\underline{t},M} := (\widetilde{X}_{\underline{t}+\underline{s}(1),M}, \ldots, \widetilde{X}_{\underline{t}+\underline{s}(m_1m_2),M})^T.$$

With the notations introduced so far we can state the following auxiliary results:

Lemma 2.13. *Let the Assumptions 1 - 4 be fulfilled with $r = 4$ and h as specified in Assumption 3. Let $\underline{c} \in \mathbb{R}^k$ be arbitrary. Then it holds:*

- $$\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2 \left|\widehat{b}_{\underline{k}}(p)\right| = \mathcal{O}_P(1), \tag{2.19}$$
- $$E^*\Big(|\varepsilon^*_{\underline{t}}|^{2w}\Big) \xrightarrow{P} E\Big(|\varepsilon_{\underline{t}}|^{2w}\Big) \quad \forall\, w \le h+2, \tag{2.20}$$
- $$\Big(X^*_{\underline{t}_1}, \ldots, X^*_{\underline{t}_d}\Big)^T \xrightarrow{d^*} \Big(\widetilde{X}_{\underline{t}_1}, \ldots, \widetilde{X}_{\underline{t}_d}\Big)^T \text{ in } P\text{-prob.} \tag{2.21}$$
  for all $d \ge 1$ and all $\underline{t}_1, \ldots, \underline{t}_d \in \mathbb{Z}^2$,
- $$E^*\Big(\Big|\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big|^{2+2/(h+1)}\Big) = \mathcal{O}_P(1),\ E\Big(\Big|\underline{c}^T g(\widetilde{\mathbf{Y}}_{\underline{t},M})\Big|^{2+2/(h+1)}\Big) \le C \tag{2.22}$$
  uniformly for all $t \in \mathbb{Z}^2$,
- $$\mathrm{Cov}^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{h},M}), \underline{c}^T g(\mathbf{Y}^*_{\underline{0},M})\Big) \xrightarrow{P} \mathrm{Cov}\Big(\underline{c}^T g(\widetilde{\mathbf{Y}}_{\underline{h},M}), \underline{c}^T g(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big) \tag{2.23}$$
  for all $\underline{h} \in \mathbb{Z}^2$.
- *The series $\Sigma^{(u,v)} := \sum_{\underline{h}\in\mathbb{Z}^2} \mathrm{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)$ converges* (2.24)
  absolutely for all $1 \le u, v \le k$.

The following auxiliary result will also be used several times:

Lemma 2.14. *Let the Assumptions 1 - 4 be fulfilled with $r = 4$. Let $W \subset \Theta \cup \{\underline{0}\}$ be any subset of vectors in the upper half-plane Θ or in the origin. We define $\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}$ and $\mathbf{Y}_{\underline{t}}^{*(W)}$ to be truncated versions of $\widetilde{\mathbf{Y}}_{\underline{t}}$ and $\mathbf{Y}_{\underline{t}}^{*}$, respectively, where*

$$\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)} := (\widetilde{X}_{\underline{t}+\underline{s}(1)}^{(W)}, \dots, \widetilde{X}_{\underline{t}+\underline{s}(m_1 m_2)}^{(W)})^T, \quad \mathbf{Y}_{\underline{t}}^{*(W)} := (X_{\underline{t}+\underline{s}(1)}^{*(W)}, \dots, X_{\underline{t}+\underline{s}(m_1 m_2)}^{*(W)})^T,$$

and

$$\widetilde{X}_{\underline{t}}^{(W)} := \sum_{\underline{k} \in W \setminus \{\underline{0}\}} b_{\underline{k}} \, \widetilde{\varepsilon}_{\underline{t}-\underline{k}} + \widetilde{\varepsilon}_{\underline{t}} \, \mathbb{1}_{\{\underline{0} \in W\}}, \quad X_{\underline{t}}^{*(W)} := \sum_{\underline{k} \in W \setminus \{\underline{0}\}} \widehat{b}_{\underline{k}}(p) \, \varepsilon_{\underline{t}-\underline{k}}^{*} + \varepsilon_{\underline{t}}^{*} \, \mathbb{1}_{\{\underline{0} \in W\}}.$$

Then there exists $C < \infty$, such that it holds for any $\underline{t} \in \mathbb{Z}^2$ and any $v = 1, \dots, k$

$$\begin{aligned} \left\| g_v(\widetilde{\mathbf{Y}}_{\underline{t}}) - g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}) \right\|_2 &\leq C \cdot \left(\sum_{\underline{k} \in \Theta \setminus W} |b_{\underline{k}}| + \mathbb{1}_{\{\underline{0} \notin W\}} \right), \\ \left\| g_v(\mathbf{Y}_{\underline{t}}^{*}) - g_v(\mathbf{Y}_{\underline{t}}^{*(W)}) \right\|_{*2} &\leq \mathcal{O}_P(1) \cdot \left(\sum_{\underline{k} \in \Theta \setminus W} \left| \widehat{b}_{\underline{k}}(p) \right| + \mathbb{1}_{\{\underline{0} \notin W\}} \right), \end{aligned}$$

where $\|z\|_2 := \left(E(z)^2\right)^{1/2}$ and $\|z\|_{*2} := \left(E^*(z)^2\right)^{1/2}$ denote the usual L^2-norms.

The previous lemma explicitly incorporates the two cases $\underline{0} \in W$ and $\underline{0} \notin W$, both of which will be needed in the proofs later on. The proofs of the lemmas from this section can be found in section 2.7, the proof of Theorem 2.12 in section 2.6.2.

2.4 Applications

In this section we will give a few examples of prominent statistics to which the check criterion derived in the previous section can be applied. These statistics are the sample mean, sample autocorrelations (for spatial processes sometimes also referred to as the sample correlogram) and the standardized sample variogram. It will be shown under appropriate conditions that the AR sieve bootstrap works asymptotically for the sample mean; for the latter two statistics it will be shown that it works for data generated by linear processes. For a simulation study concerning sample autocorrelations, see section 2.5.

Example 2.15. (Sample mean) We can use the AR sieve bootstrap procedure for the sample mean, even for processes which are not centered as required per

Assumption 1. Let $(Z_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a strictly stationary process with mean μ which, other than being non-centered, fulfils the conditions stated in Assumption 1. Since all autocovariances of $(Z_{\underline{t}})$ and the centered process $(X_{\underline{t}}) := (Z_{\underline{t}} - \mu)$ coincide, $(X_{\underline{t}})$ obviously fulfils Assumption 1. Now let $\{Z_{\underline{t}},\, \underline{t} \in \Pi\}$ be a data sample generated by $(Z_{\underline{t}})$. We apply the bootstrap procedure described in section 2.1 to the data $\{Z_{\underline{t}},\, \underline{t} \in \Pi\}$, which produces bootstrap samples $\{X^*_{\underline{t}},\, \underline{t} \in \Pi\}$, generated by

$$X^*_{\underline{t}} = \sum_{\underline{k}\in\Theta(p)} \widehat{a}_{\underline{k}}(p)\, X^*_{\underline{t}-\underline{k}} + \varepsilon^*_{\underline{t}}.$$

Then, compute $Z^*_{\underline{t}} := \overline{Z} + X^*_{\underline{t}}$ for all $\underline{t} \in \Pi$, where $\overline{Z} := |\Pi|^{-1} \sum_{\underline{t}\in\Pi} Z_{\underline{t}}$ (for the bootstrap data, $\overline{Z}^*$ is analogously defined). We can approximate the distribution of $n(\overline{Z} - \mu)$ by the one of $n(\overline{Z}^* - \overline{Z})$. Asymptotic validity of this approach can be established via Theorem 2.12 in the following way:
The companion process associated with $(X_{\underline{t}})$ is denoted by $(\widetilde{X_{\underline{t}}})$ and we define $\widetilde{Z}_{\underline{t}} := \widetilde{X}_{\underline{t}} + \mu$. The functions f and g in assumption 3 can be chosen appropriately such that T_n is the sample mean of $\{X_{\underline{t}},\, \underline{t} \in \Pi\}$, and $\widetilde{T}_n = \overline{\widetilde{X}}$ is the mean of $\{\widetilde{X}_{\underline{t}},\, \underline{t} \in \Pi\}$. For the linear process $(\widetilde{Z}_{\underline{t}})$, with an obvious notation for $\overline{\widetilde{Z}}$, it is known that

$$n\left(\overline{\widetilde{Z}} - \mu\right) = n\left(\overline{\widetilde{X}}\right) = n\,\widetilde{T}_n \xrightarrow{d} \mathcal{N}\Big(0\,,\, \textstyle\sum_{\underline{h}\in\mathbb{Z}^2} \gamma_{\widetilde{Z}}(\underline{h})\Big),$$

where $\gamma_{\widetilde{Z}}$ denotes the autocoavariance function of $(\widetilde{Z}_{\underline{t}})$. Noting that $\overline{Z}^* = \overline{Z} + \overline{X}^*$, it follows immediately from Theorem 2.12

$$n\left(\overline{Z}^* - \overline{Z}\right) = n\left(\overline{X}^*\right) = n\,T^*_n \xrightarrow{d^*} \mathcal{N}\Big(0\,,\, \textstyle\sum_{\underline{h}\in\mathbb{Z}^2} \gamma_{\widetilde{Z}}(\underline{h})\Big) \quad \text{in prob.} \tag{2.25}$$

For the sample mean $\overline{Z}$ of the actually observed data it holds under suitable regularity conditions that

$$n\left(\overline{Z} - \mu\right) \xrightarrow{d} \mathcal{N}\Big(0\,,\, \textstyle\sum_{\underline{h}\in\mathbb{Z}^2} \gamma_Z(\underline{h})\Big). \tag{2.26}$$

Now observe that $(Z_{\underline{t}})$ and $(\widetilde{Z}_{\underline{t}})$ have identical second order properties per definition. In particular, $\gamma_Z(\underline{h}) = \gamma_{\widetilde{Z}}(\underline{h})$ for all lags $\underline{h} \in \mathbb{Z}^2$. Thus, the limiting distributions in (2.25) and (2.26) coincide and it follows

$$d_K\left(\mathcal{L}^*\Big(n\left(\overline{Z}^* - \overline{Z}\right)\Big), \mathcal{L}\Big(n\left(\overline{Z} - \mu\right)\Big)\right) = o_P(1).$$

Therefore, the AR sieve bootstrap proposal is asymptotically valid for the sample mean under the stated conditions. □

In contrast to the preceeding example, the limiting distribution of sample autocovariances does not depend exclusively on second-order properties of the underlying process. This result is well-known, particularly for the time-series case, i.e. $d = 1$. Even if the data are generated by a linear spatial process, that is a process of the form

$$X_{\underline{t}} = \sum_{\underline{\nu} \in \mathbb{Z}^2} \alpha_{\underline{\nu}} \, u_{\underline{t}-\underline{\nu}}, \tag{2.27}$$

with absolutely summable coefficients $(\alpha_{\underline{\nu}})_{\underline{\nu} \in \mathbb{Z}^2}$ and an i.i.d. white noise process $(u_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ with finite fourth moments, the limiting variance depends on the fourth-order cumulants of $(u_{\underline{t}})$. This can be verified with analogous calculations as for the times series case, cf. Brockwell and Davis (1991), Proposition 7.3.4. However, the situation is different if one switches to sample autocorrelations of linear processes, instead of autocovariances. Then, the limiting distribution depends only on the autocorrelations of the underlying process, as shows the following theorem, which is a direct generalisation of the well-known Bartlett formula for time series, cf. Brockwell and Davis (1991), Proposition 7.2.1.:

Lemma 2.16. *Let $(X_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ be a linear spatial process as defined in (2.27), i.e. with i.i.d. white noise and finite fourth moments, and with autocorrelation function ρ. For the sample autocorrelations $\widehat{\rho}(\underline{h}) = \widehat{\gamma}(\underline{h})/\widehat{\gamma}(\underline{0})$, with $\widehat{\gamma}(\cdot)$ as defined in (2.1), we define the comparative quantity $\check{\rho}(\underline{h}) := \check{\gamma}(\underline{h})/\check{\gamma}(\underline{0})$ with*

$$\check{\gamma}(\underline{h}) := \frac{1}{|\Pi|} \sum_{\underline{t} \in \Pi} X_{\underline{t}+\underline{h}} \, X_{\underline{t}},$$

where $\Pi = \{\underline{t} \in \mathbb{Z}^2 : \; 1 \leq t_1, t_2 \leq n\}$. $\check{\rho}(\underline{h})$ and $\widehat{\rho}(\underline{h})$ are asymptotically equivalent. Then it holds

$$n^2 \operatorname{Cov}(\check{\rho}(\underline{h}), \check{\rho}(\underline{k})) \longrightarrow V(\underline{h}, \underline{k}), \qquad \text{as } n \to \infty,$$

where

$$\begin{aligned} V(\underline{h}, \underline{k}) \;\; &= \;\; \sum_{\underline{r} \in \mathbb{Z}^2} \Big\{ 2\rho(\underline{r})^2 \rho(\underline{k})\rho(\underline{h}) - 2\rho(\underline{r}+\underline{k})\rho(\underline{r})\rho(\underline{h}) - 2\rho(\underline{r}-\underline{h})\rho(\underline{r})\rho(\underline{k}) \\ &\qquad + \rho(\underline{r}-\underline{h}+\underline{k})\rho(\underline{r}) + \rho(\underline{r}+\underline{k})\rho(\underline{r}-\underline{h}) \Big\}. \end{aligned}$$

The proof is analogous to the time-series case and can be found in section 2.7.

Example 2.17. (Sample autocorrelations/correlogram) Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a spatial process fulfilling Assumption 1 with corresponding companion process $(\widetilde{X}_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$. We consider the autocorrelation function $\rho(\underline{h}) = \gamma(\underline{h})/\gamma(\underline{0})$ at lag $\underline{h}$, together with the usual estimator $T_n := \widehat{\rho}(\underline{h}) = \widehat{\gamma}(\underline{h})/\widehat{\gamma}(\underline{0})$, where $\widehat{\gamma}(\cdot)$ is given by (2.1). For spatial processes, $\rho(\underline{h})$ (and accordingly $\widehat{\rho}(\underline{h})$) are often referred to as the (sample) correlogram, cf. Cressie (1993), Section 2.3.2. $\widetilde{T}_n$ denotes the same estimator as T_n, but applied to a sample from the companion process. Note that the autocorrelations of $(\widetilde{X}_{\underline{t}})$ are given by the function ρ as well, because all second order properties of $(\widetilde{X}_{\underline{t}})$ and $(X_{\underline{t}})$ coincide. Under suitable assumptions on the dependence structure of the process, such as weak dependence or mixing conditions, it is known that

$$n(\widehat{\rho}(\underline{h}) - \rho(\underline{h})) \xrightarrow{d} \mathcal{N}(0, \tau_X^2), \quad n(\widetilde{T}_n - \rho(\underline{h})) \xrightarrow{d} \mathcal{N}(0, \tau_{\widetilde{X}}^2),$$

where the limiting variances τ_X^2 and $\tau_{\widetilde{X}}^2$ in general depend on the fourth order cumulants of $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$, respectively. Hence, it follows $\tau_X^2 \neq \tau_{\widetilde{X}}^2$ in general, because $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$ share second order but not fourth order properties. For T_n^*, denoting the sample autocorrelation applied to the bootstrap sample $\{X_{\underline{t}}^*, \underline{t} \in \Pi\}$, Theorem 2.12 yields

$$n\Big(T_n^* - f(\underline{\theta}^*)\Big) \xrightarrow{d} \mathcal{N}(0, \tau_{\widetilde{X}}^2).$$

Therefore, $\tau_X^2 \neq \tau_{\widetilde{X}}^2$ implies that the AR sieve bootstrap in general is asymptotically not valid for sample autocorrelations.
However, if the data are generated by a linear process $(X_{\underline{t}})$ as given by (2.27), Lemma 2.16 shows that the limiting variance of $n(\check{\rho}(\underline{h}) - \rho(\underline{h}))$ is given by

$$\begin{aligned} \tau_X^2 \;=\; \sum_{\underline{r}\in\mathbb{Z}^2} \Big\{ & 2\rho(\underline{r})^2\rho(\underline{h})^2 - 2\rho(\underline{r}+\underline{h})\rho(\underline{r})\rho(\underline{h}) - 2\rho(\underline{r}-\underline{h})\rho(\underline{r})\rho(\underline{h}) \\ & + \rho(\underline{r})^2 + \rho(\underline{r}+\underline{h})\rho(\underline{r}-\underline{h}) \Big\}. \end{aligned} \tag{2.28}$$

Since $\check{\rho}(\underline{h})$ and $\widehat{\rho}(\underline{h})$ are asymptotically equivalent, $n(\widehat{\rho}(\underline{h}) - \rho(\underline{h}))$ also has limiting variance τ_X^2. This expression depends only on the autocorrelations of the underlying process, which coincide for $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$. Thus, it follows for this case $\tau_X^2 = \tau_{\widetilde{X}}^2$, and the bootstrap procedure is asymptotically valid for sample autocorrelations of data generated from linear processes. □

Remark 2.18. When checking for asymptotic validity of the AR sieve bootstrap procedure, it is of critical importance to ensure that the limiting distributions of T_n and $\widetilde{T}_n$ are identical, as has been done in the previous examples. In general, this will be the case whenever the limiting distribution depends only on second order entities such as autocovariances or the spectral density of the underlying process. For data generated by a linear process $X_{\underline{t}} = \sum_{\underline{\nu}\in\mathbb{Z}^2} \alpha_{\underline{\nu}}\, u_{\underline{t}-\underline{\nu}}$, one might be tempted to conclude that $(X_{\underline{t}})$ and its companion process $(\widetilde{X_{\underline{t}}})$ are identical since $(u_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ is already i.i.d.. However, Example 3.2 from Kreiss, Paparoditis and Politis (2011) shows for the special case of time series that this is not the case. To be precise, the companion process $(\widetilde{X_{\underline{t}}})$ is always derived from the AR representation (2.5), where $(\varepsilon_{\underline{t}})$ is the uniquely determined innovation process of $(X_{\underline{t}})$. Even if the process has linear representation $X_{\underline{t}} = \sum_{\underline{\nu}\in\mathbb{Z}^2} \alpha_{\underline{\nu}}\, u_{\underline{t}-\underline{\nu}}$ with i.i.d. noise $(u_{\underline{t}})$, its innovation process might differ from $(u_{\underline{t}})$, and might be only uncorrelated but not i.i.d.. Remark 2.1 of Kreiss, Paparoditis and Politis (2011) gives a specific example of this situation. Therefore, linear processes are in general not identical to their companion processes, which makes a careful inspection of the limiting distributions as in the previous examples a necessity. □

Example 2.19. (Standardized sample variogram) Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a spatial process fulfilling Assumption 1 with autocovariance function γ. The variogram at lag $\underline{h} \in \mathbb{Z}^2$ is defined as

$$V(\underline{h}) = \mathrm{Var}(X_{\underline{t}} - X_{\underline{t}+\underline{h}}) = E\Big((X_{\underline{t}} - X_{\underline{t}+\underline{h}})^2\Big) = 2\gamma(\underline{0}) - 2\gamma(\underline{h})$$

for centered fields, and $V^{(s)}(\underline{h}) := V(\underline{h})/\gamma(\underline{0})$ is called the standardized variogram. Using the notation from (2.1), two classical estimators for $V(\underline{h})$ are given by

$$\widehat{V}_1(\underline{h}) = 2\widehat{\gamma}(\underline{0}) - 2\widehat{\gamma}(\underline{h}), \quad \widehat{V}_2(\underline{h}) = \frac{1}{|\Pi_{\underline{h}}|}\sum_{\underline{t}\in\Pi_{\underline{h}}} (X_{\underline{t}} - X_{\underline{t}+\underline{h}})^2,$$

which are asymptotically equivalent, cf. Cressie (1993), Section 2.4. In particular, one can easily check that

$$n\Big(\widehat{V}_1(\underline{h}) - \widehat{V}_2(\underline{h})\Big) = o_P(1). \tag{2.29}$$

Versions of both of these estimators are included in the class of functions of generalized means, as given by Assumption 3. Furthermore, both $\widehat{V}_1(\underline{h})$ and $\widehat{V}_2(\underline{h})$ can be used to construct standardized sample variogram estimators via $\widehat{V}_j^{(s)}(\underline{h}) := \widehat{V}_j(\underline{h})/\widehat{\gamma}(\underline{0})$, $j = 1, 2$. It holds

$$\widehat{V}_1^{(s)}(\underline{h}) = 2 - 2\widehat{\rho}(\underline{h}).$$

Now assume the data are generated by a linear process. Then it follows from Example 2.17, with τ_X^2 as defined there,

$$n\Big(\widehat{V}_1^{(s)}(\underline{h}) - V^{(s)}(\underline{h})\Big) = (-2)\cdot n\Big(\widehat{\rho}(\underline{h}) - \rho(\underline{h})\Big) \xrightarrow{d} \mathcal{N}(0, 4\tau_X^2),$$

and $n\Big(\widehat{V}_2^{(s)}(\underline{h}) - V^{(s)}(\underline{h})\Big)$ has the very same limiting distribution due to (2.29). An analogous argumentation as in Example 2.17 therefore yields asymptotic validity of the AR sieve bootstrap procedure for the standardized sample variogram, as long as the data are generated by a linear spatial process. □

Remark 2.20. Our main result Theorem 2.12 provides a check criterion for asymptotic validity of the AR sieve bootstrap for all statistics from Assumption 3. This class of statistics contains, among other things, the statistics from Examples 2.15-2.19. However, we conjecture that analogous results can be proven, in the same spirit as in the proof of Theorem 2.12, for a much wider class of statistics beyond those covered by Assumption 3. If T_n denotes an estimator for some parameter θ, under the condition that $\mathcal{L}\Big(c_n(T_n - \theta)\Big)$ has a non-degenerated limiting distribution for some sequence (c_n), we conjecture that the AR sieve bootstrap procedure is asymptotically valid, as long as the limiting distribution depends on second order properties of the underlying process, only.
For example, according to Section 4.5 in Guyon (1995), one can prove central limit theorems for kernel-based nonparametric spectral density estimators for strictly stationary spatial processes under appropriate mixing conditions. The limiting distribution then depends exclusively on the spectral density of the underlying process, which is a second order quantity, and we conjecture that the AR sieve bootstrap is asymptotically valid in this situation. □

2.5 A simulation study

In this section, we will present simulation results that compare the performance of the AR sieve bootstrap to classic normal approximations and block bootstrap methods. We generated square-shaped samples $\{X_{\underline{t}} = X_{t_1,t_2} : 1 \leq t_1, t_2 \leq n\}$ as defined in section 2.1, where the sample size is set to be $n = 15$ which corresponds to $15 \times 15 = 225$ observations. The samples are generated by a moving average model given by

$$X_{t_1,t_2} = e_{t_1,t_2} + 0.5 \cdot e_{t_1+1,t_2} - 0.2 \cdot e_{t_1-1,t_2} + 0.3 \cdot e_{t_1,t_2+1} + 0.1 \cdot e_{t_1,t_2-1}, \quad (2.30)$$

where $(e_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ is an i.i.d. white noise process with marginal distribution $\mathcal{N}(0,1)$. The process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ fulfils the conditions of Assumption 1. Furthermore, each realisation $X_{\underline{t}}$ depends on noise terms from four different directions, two from the lower and two from the upper half-plane, cf. section 2.1. This means that the process is *not* 'tailor-made' for an AR approximation in the direction of the lower half-plane as performed in the AR sieve algorithm. If we had chosen an underlying process generated exclusively from noise variables in the lower half-plane, the AR sieve bootstrap would clearly have had an advantage compared to the other methods. However, the data generating process from (2.30) does not 'favor' any direction of one-sided autoregressive fits; one could as well fit models that are one-sided with respect to the upper, left or right half-plane.

The statistic that we investigated is the sample autocorrelation $\widehat{\rho}(\underline{h})$ as defined in Example 2.17, with $\underline{h} = (1, -1)^T$. For the process from (2.30), the true autocorrelation is given by $\rho(1, -1) = 0.13/1.39$. We approximated the distribution of

$$n\Big(\widehat{\rho}(1, -1) - \rho(1, -1)\Big) \tag{2.31}$$

for $n = 15$ with a normal approximation and with the AR sieve bootstrap, via the empirical distribution of $n(\widehat{\rho}^*(1, -1) - \widehat{\rho}(1, -1))$. To implement the normal approximation, we considered the limiting distribution of (2.31) given by $\mathcal{N}(0, \tau_X^2)$ with τ_X^2 from (2.28), cf. Example 2.17. For the process $(X_{\underline{t}})$ from (2.30) one can easily verify that τ_X^2 is given by

$$\tau_X^2 = \sum_{|\underline{r}|_\infty \leq 2} \Big\{2\rho(\underline{r})^2 \rho(1, -1)^2 - \ldots\Big\}, \tag{2.32}$$

since all summands with $|\underline{r}|_\infty := \max\{|r_1|, |r_2|\} > 2$ vanish due to $\rho(\underline{r}) = 0$ for all $|\underline{r}|_\infty > 2$. Hence, we estimated τ_X^2 by replacing ρ with $\widehat{\rho}$ in (2.32). It should be noted that this approach represents a best-case szenario for the normal approximation because we used the additional information that for the present data τ_X^2 has the special form (2.32), i.e. we chose the optimal point of cutting off the infinite sum in (2.28). For real-world data, this information would not be known, and one would have to estimate τ_X^2 based on equation (2.28) by cutting off the infinite sum at some non-optimal point which would generate an additional error in the estimation.

Figure 2.3 shows the comparison of three different choices for the order p of the AR sieve bootstrap. We simulated the 95%-quantile of the distribution of (2.31) for $n = 15$. In each iteration, we generated $M = 500$ bootstrap samples to approximate this quantile, subsequently using the AR sieve bootstrap with orders $p = 1$, $p = 2$ and $p = 3$. We also calculated the normal approximation estimate of the quantile in each iteration as described previously. All of this was carried out for $N = 50$

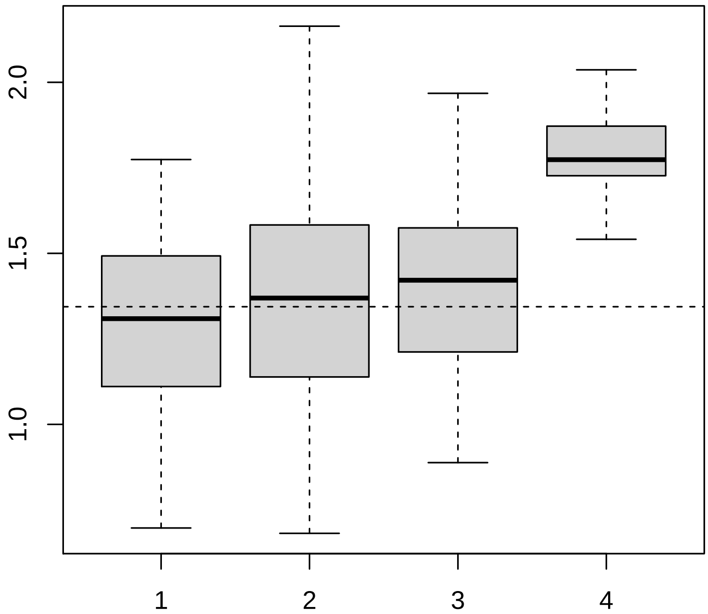

Figure 2.3: Approximations of the 95%-quantile of the distribution of $n(\widehat{\rho}(1,-1)-\rho(1,-1))$ for $n=15$; boxplots based on $N=50$ iterations. From left to right: AR sieve bootstrap (based on $M=500$ repetitions) with $p=1$, $p=2$ and $p=3$, followed by the normal approximation at the right end. Target value given by the horizontal dashed line.

independent iterations to generate boxplots of the locations of the estimates. In Figure 2.3, the three AR sieve approximations are shown in the three boxplots to the left (from left to right: $p = 1, 2, 3$) and the normal approximation values are given by the boxplot to the right end. The target value, i.e. the 95%-quantile of the distribution of (2.31), is determined from Monte-Carlo simulations with 500000 repetitions and illustrated by the horizontal dashed line. One can see that the AR sieve bootstrap works very well compared to the normal approximation, even for small orders p and even though we have used an optimal normal approximation as described earlier. Of course, we have to concede that for a moving average model such as (2.30) the autoregressive coefficients will decay quickly and, thus, the dependence structure within the data can be depicted well with AR models of small order.

We also compared the performances of the AR sieve bootstrap and block bootstrap techniques. The parameters were chosen identically to the previous simulation, i.e. $N = 50$ iterations of the AR sieve bootstrap and the moving block bootstrap with various block lengths (each based on $M = 500$ repetitions) were carried out. The target was again the 95%-quantile of the distribution of (2.31) for $n = 15$. The order

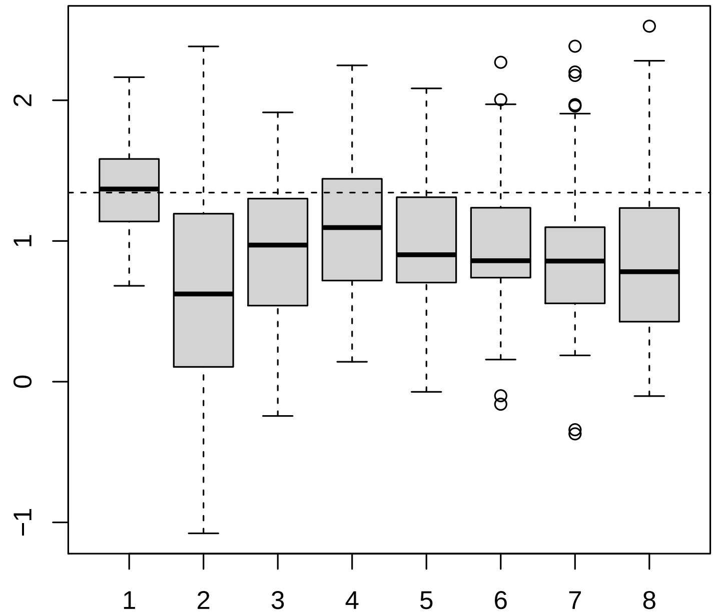

Figure 2.4: Approximations of the 95%-quantile of the distribution of $n(\widehat{\rho}(1,-1)-\rho(1,-1))$ for $n = 15$; boxplots based on $N = 50$ iterations and each bootstrap method based on $M = 500$ repetitions. From left to right: AR sieve bootstrap with $p = 2$ in column 1, followed by the block bootstrap with block length $l = 2, \ldots, 8$ (block length l depicted in column l). Target value given by the horizontal dashed line.

of the AR sieve bootstrap was fixed to $p = 2$ and we considered block lengths of $l = 2, \ldots, 8$. Here, the block length refers to square-shaped blocks, i.e. a block length of l means drawing blocks of $l \times l$ observations from the original data sample and then sticking the blocks together to form a sample of size $n \times n$. The result can be seen in Figure 2.4. The boxplot to the left corresponds to the AR sieve bootstrap (column 1) and the results for the block bootstrap are given in columns $2, \ldots, 8$ with block length l depicted in column l. Arguably the best result for the block bootstrap is achieved for $l = 4$; however, the AR sieve bootstrap performs considerably stronger than all block bootstrap approaches implemented here.

In order to show that the results obtained so far are not only specific to the 95%-quantile but to the distribution of (2.31) as a whole, we will now look at an approximation of the variance of this distribution instead of a single quantile. Figure 2.5 shows these approximations of the variance with all parameters as before, i.e. $n = 15$, $N = 50$, $M = 500$ and $p = 2$. The AR sieve bootstrap is depicted in column 1, the block bootstrap in columns 2 and 3 (block lengths $l = 5, 6$) and the normal

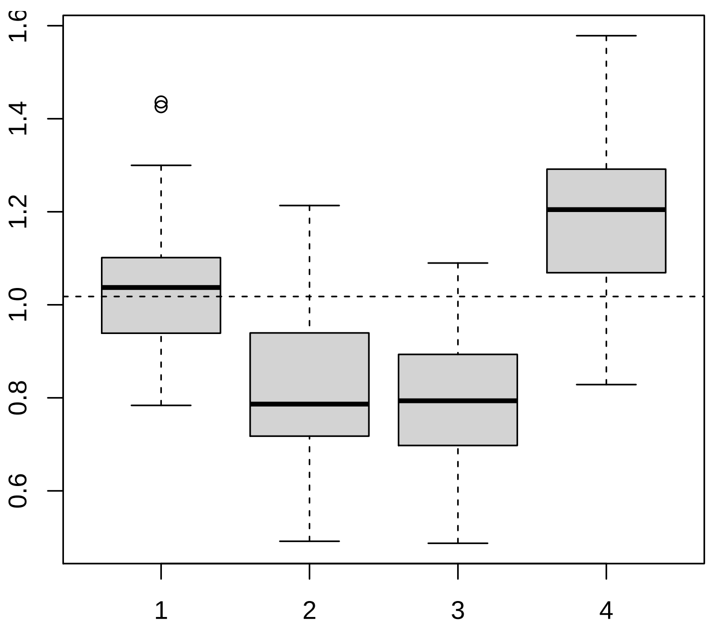

Figure 2.5: Approximations of the variance of the distribution of $n(\widehat{\rho}(1,-1)-\rho(1,-1))$ for $n=15$; boxplots based on $N=50$ iterations and each bootstrap method based on $M=500$ repetitions. From left to right: AR sieve bootstrap with $p=2$ in column 1, followed by the block bootstrap with block length $l=5,6$ (in columns 2 and 3), normal approximation in column 4. Target value given by the horizontal dashed line.

approximation in column 4. Similar to what happens for the 95%-quantile, the AR sieve bootstrap outperforms the other methods.

To conclude this section, we modified some of the parameters from the simulations performed so far. The data are still generated by a moving average model, but now following the model equation

$$X_{t_1,t_2} = e_{t_1,t_2} + 4 \cdot e_{t_1+1,t_2} - 5 \cdot e_{t_1-1,t_2} + 3 \cdot e_{t_1,t_2+1} - 2 \cdot e_{t_1,t_2-1}, \qquad (2.33)$$

where the noise is no longer symmetricly distributed but has an i.i.d. centered exponential distribution. In this model, the dependence of neighbouring random variables is higher than in model (2.30). For example, the true autocorrelation at lag $\underline{h} = (1,-1)^T$ is here given by $\rho(1,-1) = 0.4$ compared to $\rho(1,-1) \approx 0.094$ in model (2.30). We also increase the sample size to $n = 25$ – corresponding to $25 \times 25 = 625$ observations – and choose the order $p = 4$ for the AR sieve bootstrap. Figure 2.6 shows the results for the approximation of the 95%-quantile of the distribution of (2.31) for $n = 25$; from left to right: AR sieve bootstrap, block bootstrap with

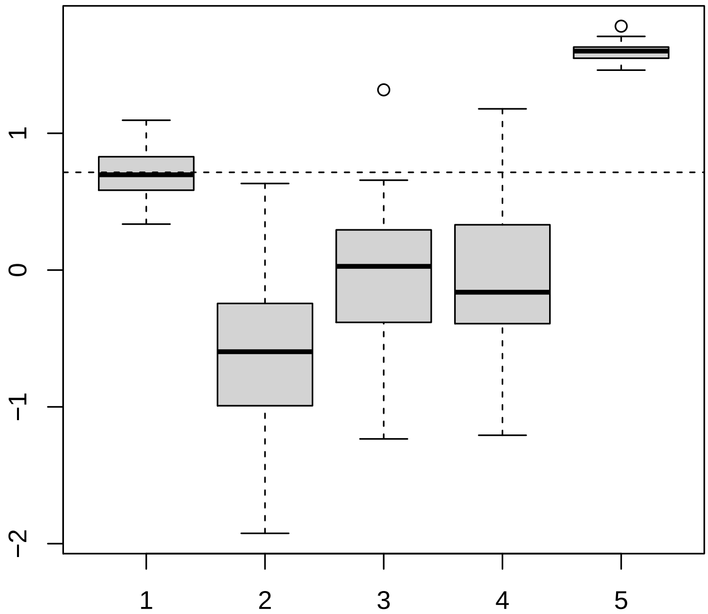

Figure 2.6: Approximations of the 95%-quantile of the distribution of $n(\widehat{\rho}(1,-1)-\rho(1,-1))$ for $n=25$; boxplots based on $N=50$ iterations and each bootstrap method based on $M=300$ repetitions. From left to right: AR sieve bootstrap with $p=4$ in column 1, followed by the block bootstrap with block length $l=8,9,10$ (in columns $2,3,4$), normal approximation in column 5. Target value given by the horizontal dashed line.

$l=8,9,10$ and the normal approximation. It can be seen that, for this increased sample size, the normal approximation is close to its limit which, however, differs considerably from the true quantile of the finite sample distribution. This is mainly due to a negative bias for the distribution of (2.31) which can be obtained from the Monte Carlo simulations that were performed to determine the 95%-quantile. The block bootstrap clearly does not show desirable results, this might stem from the increased dependence between neighbouring realisations in the present model compared to the model used previously. However, the AR sieve bootstrap performs very well for this choice of (increased values of) n and p. This emphasizes the fact that convergence of the AR sieve bootstrap can be achieved as long as $p=p(n)\to\infty$ at an appropriate rate, for $n\to\infty$, as was proven in the previous sections.

2.6 Proofs of the main results

2.6.1 Proof of Theorem 2.7

In order to write the Yule-Walker equations (2.9) in compact form we denoted $\bar{p} = 2p(p+1)$ and introduced the arbitrary but fixed enumeration $\underline{k}_1, \ldots, \underline{k}_{\bar{p}}$ of the vectors $\underline{k} \in \Theta(p)$. Now we extend this enumeration to the infinite but countable set Θ, by choosing an arbitrary enumeration $\underline{k}_{\bar{p}+1}, \underline{k}_{\bar{p}+2}, \ldots$ of the vectors $\underline{k} \in \Theta \setminus \Theta(p)$ such that

$$\Theta = \{\underline{k}_1, \ldots, \underline{k}_{\bar{p}}\} \cup \{\underline{k}_{\bar{p}+1}, \underline{k}_{\bar{p}+2}, \ldots\}.$$

While the finite predictor coefficients $(a_{\underline{k}}(p))_{\underline{k}\in\Theta(p)}$ are given by (2.9), Lemma 2.1 shows that the autoregressive coefficients $(a_{\underline{k}})_{\underline{k}\in\Theta}$ determine the L^2-projection of $X_{\underline{t}}$ onto $\overline{sp}\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta\}$. Therefore, $X_{\underline{t}} - \sum_{\underline{k}\in\Theta} a_{\underline{k}} X_{\underline{t}-\underline{k}}$ is orthogonal to each $X_{\underline{t}-\underline{s}}$, $\underline{s} \in \Theta$. Equivalently, with the introduced enumeration of Θ this means

$$\mathrm{Cov}\Big(X_{\underline{t}} - \sum_{j=1}^{\infty} a_{\underline{k}_j} X_{\underline{t}-\underline{k}_j}, X_{\underline{t}-\underline{k}_m}\Big) = \gamma(\underline{k}_m) - \sum_{j=1}^{\infty} a_{\underline{k}_j} \gamma(\underline{k}_m - \underline{k}_j) = 0 \quad \forall\, m \in \mathbb{N}.$$

From this system of equations we consider only those ones with $m = 1, \ldots, \bar{p}$, which is equivalent to

$$\Gamma(p) \cdot \begin{pmatrix} a_{\underline{k}_1} \\ \vdots \\ a_{\underline{k}_{\bar{p}}} \end{pmatrix} + \sum_{j=\bar{p}+1}^{\infty} a_{\underline{k}_j} \begin{pmatrix} \gamma(\underline{k}_1 - \underline{k}_j) \\ \vdots \\ \gamma(\underline{k}_{\bar{p}} - \underline{k}_j) \end{pmatrix} = \begin{pmatrix} \gamma(\underline{k}_1) \\ \vdots \\ \gamma(\underline{k}_{\bar{p}}) \end{pmatrix}.$$

Since the right-hand sides of this system and (2.9) coincide, we can infer

$$\begin{pmatrix} a_{\underline{k}_1}(p) - a_{\underline{k}_1} \\ \vdots \\ a_{\underline{k}_{\bar{p}}}(p) - a_{\underline{k}_{\bar{p}}} \end{pmatrix} = \Gamma(p)^{-1} \cdot \sum_{j=\bar{p}+1}^{\infty} a_{\underline{k}_j} \begin{pmatrix} \gamma(\underline{k}_1 - \underline{k}_j) \\ \vdots \\ \gamma(\underline{k}_{\bar{p}} - \underline{k}_j) \end{pmatrix}. \tag{2.34}$$

In the following we will denote the (n,r)-th entry of $\Gamma(p)^{-1}$ by $(\Gamma(p)^{-1})^{(n,r)}$. We are interested in a weighted sum of the absolute values of the entries on the left-hand side of (2.34). For $s \in \mathbb{N}_0$ such that $s + 1 < r$ we get

$$\begin{aligned}
&\sum_{n=1}^{\bar{p}} (1 + |\underline{k}_n|_\infty)^s \, |a_{\underline{k}_n}(p) - a_{\underline{k}_n}| \\
=\; &\sum_{n=1}^{\bar{p}} (1 + |\underline{k}_n|_\infty)^s \Big| \sum_{j=\bar{p}+1}^{\infty} a_{\underline{k}_j} \sum_{r=1}^{\bar{p}} (\Gamma(p)^{-1})^{(n,r)} \gamma(\underline{k}_r - \underline{k}_j) \Big| \\
\le\; &\sum_{j=\bar{p}+1}^{\infty} |a_{\underline{k}_j}| \sum_{r=1}^{\bar{p}} |\gamma(\underline{k}_r - \underline{k}_j)| \max_{r=1,\ldots,\bar{p}} \sum_{n=1}^{\bar{p}} (1 + |\underline{k}_n|_\infty)^s \left|(\Gamma(p)^{-1})^{(n,r)}\right|
\end{aligned} \tag{2.35}$$

We denote the max-column-sum norm of an arbitrary $n \times n$-matrix B by $\|B\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^{n} |B^{(i,j)}|$. It is well-known that $\|\cdot\|_1$ is submultiplicative which allows us to derive

$$\begin{aligned} & \max_{r=1,\dots,\bar{p}} \sum_{n=1}^{\bar{p}} (1+|\underline{k}_n|_\infty)^s \left|(\Gamma(p)^{-1})^{(n,r)}\right| \\ = & \left\| \operatorname{diag}\left[(1+|\underline{k}_1|_\infty)^s, \dots, (1+|\underline{k}_{\bar{p}}|_\infty)^s\right] \cdot \Gamma(p)^{-1} \right\|_1 \\ \leq & \max_{n=1,\dots,\bar{p}} (1+|\underline{k}_n|_\infty)^s \cdot \left\|\Gamma(p)^{-1}\right\|_1 . \end{aligned}$$

Hence, (2.35) can be bounded from above by

$$\begin{aligned} & \left\|\Gamma(p)^{-1}\right\|_1 \cdot \sum_{j=\bar{p}+1}^{\infty} \max_{n=1,\dots,\bar{p}} (1+|\underline{k}_n|_\infty)^s \, |a_{\underline{k}_j}| \sum_{r=1}^{\bar{p}} |\gamma(\underline{k}_r - \underline{k}_j)| \\ \leq & \left\|\Gamma(p)^{-1}\right\|_1 \sum_{\underline{k}\in\mathbb{Z}^2} |\gamma(\underline{k})| \cdot \sum_{j=\bar{p}+1}^{\infty} \max_{n=1,\dots,\bar{p}} (1+|\underline{k}_n|_\infty)^s \, |a_{\underline{k}_j}| . \end{aligned}$$

Since our numeration was chosen such that $\Theta(p) = \{\underline{k}_1, \dots, \underline{k}_{\bar{p}}\}$ and $\Theta \setminus \Theta(p) = \{\underline{k}_{\bar{p}+1}, \underline{k}_{\bar{p}+2}, \dots\}$, the inequality derived so far reads

$$\begin{aligned} & \sum_{\underline{k}\in\Theta(p)} (1+|\underline{k}|_\infty)^s \, |a_{\underline{k}}(p) - a_{\underline{k}}| \\ \leq & \left\|\Gamma(p)^{-1}\right\|_1 \sum_{\underline{k}\in\mathbb{Z}^2} |\gamma(\underline{k})| \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} \max_{\underline{v}\in\Theta(p)} (1+|\underline{v}|_\infty)^s \, |a_{\underline{k}}| . \end{aligned} \tag{2.36}$$

Per definition of $\Theta(p)$ we have

$$\max_{\underline{v}\in\Theta(p)} (1+|\underline{v}|_\infty)^s = (1+p)^s \leq (1+|\underline{k}|_\infty)^s \quad \forall\, \underline{k} \in \Theta \setminus \Theta(p), \tag{2.37}$$

as $|\underline{k}|_\infty \geq p+1$ for all $\underline{k} \in \Theta \setminus \Theta(p)$; this is why we need a weight function strictly nondecreasing in $|\underline{k}|_\infty$. Furthermore, it holds $\|A\|_1 \leq \sqrt{n}\,\|A\|_{\text{spec}}$ for all $n \times n$-matrices A, i.e. $\|\Gamma(p)^{-1}\|_1 \leq \sqrt{2p(p+1)}\,\|\Gamma(p)^{-1}\|_{\text{spec}}$ and

$$\sqrt{2p(p+1)} \leq \sqrt{2}\,(p+1) < \sqrt{2}\,(1+|\underline{k}|_\infty) \quad \forall\, \underline{k} \in \Theta \setminus \Theta(p).$$

Therefore, and due to (2.37) and Lemma 2.6, (2.36) can be bounded by

$$\begin{aligned} & \sqrt{2p(p+1)}\,\left\|\Gamma(p)^{-1}\right\|_{\text{spec}} \sum_{\underline{k}\in\mathbb{Z}^2} |\gamma(\underline{k})| \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1+|\underline{k}|_\infty)^s \, |a_{\underline{k}}| \\ \leq & \frac{1}{2\sqrt{2}\,\pi^2 c} \sum_{\underline{k}\in\mathbb{Z}^2} |\gamma(\underline{k})| \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1+|\underline{k}|_\infty)^{s+1} \, |a_{\underline{k}}| , \end{aligned}$$

which completes the proof. □

2.6.2 Proof of Theorem 2.12

The basic structure of this proof resembles the proof of Theorem 3.3 in Bühlmann (1997). At first, we will neglect the outer function f in T_n^* and show for the bootstrap quantities

$$(\bar{n}_1\bar{n}_2)^{-1/2}\sum_{t_1=1}^{\bar{n}_1}\sum_{t_2=1}^{\bar{n}_2}\Big(g(\mathbf{Y}^*_{\underline{t}}) - E^*\Big(g(\mathbf{Y}^*_{\underline{t}})\Big)\Big) \xrightarrow{d^*} \mathcal{N}(\underline{0},\Sigma) \text{ in prob.,} \tag{2.38}$$

where the entries of Σ are given by $\Sigma^{(u,v)} := \sum_{\underline{h}\in\mathbb{Z}^2}\operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)$, for $u, v = 1,\dots,k$. Note that (2.24) guarantees that this expression is well-defined. Since the companion process $(\widetilde{X}_{\underline{t}})$, just as the bootstrap process, is a linear spatial process (recall that the innovations $(\widetilde{\varepsilon}_{\underline{t}})$ are i.i.d.), one can follow the exact same arguments as in proving (2.38) to derive

$$(\bar{n}_1\bar{n}_2)^{-1/2}\sum_{t_1=1}^{\bar{n}_1}\sum_{t_2=1}^{\bar{n}_2}\Big(g(\widetilde{\mathbf{Y}}_{\underline{t}}) - E\Big(g(\widetilde{\mathbf{Y}}_{\underline{t}})\Big)\Big) \xrightarrow{d} \mathcal{N}(\underline{0},\Sigma) \tag{2.39}$$

with the very same limiting distribution as above. We will therefore restrict ourselves to providing a thorough reasoning of (2.38) and omit the proof of the CLT for the companion process. In the end we will incorporate the function f by applying the delta method to both CLT's which will complete the proof of Theorem 2.12 since $(\bar{n}_1\bar{n}_2)^{1/2}$ and n are asymptotically equivalent.

The strategy for proving (2.38) is the following: Firstly, one can observe that (2.38) follows if we can show

$$\frac{1}{n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\Big(g(\mathbf{Y}^*_{\underline{t}}) - E^*\Big(g(\mathbf{Y}^*_{\underline{t}})\Big)\Big) \xrightarrow{d^*} \mathcal{N}(\underline{0},\Sigma) \text{ in prob.,} \tag{2.40}$$

since the expressions in both assertions are asymptotically equivalent per definition of $\bar{n}_1, \bar{n}_2$. We will invoke the Cramér-Wold device and, in the first step, consider the truncated quantity $(\mathbf{Y}^*_{\underline{t},M})$ based on the truncated process $(X^*_{\underline{t},M})$ introduced in (2.18). For arbitrary $M \in \mathbb{N}$ and for all $\underline{c} \in \mathbb{R}^k$ we will show

$$\frac{1}{n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big) \xrightarrow{d^*} \mathcal{N}\Big(0, \underline{c}^T\Sigma_M\,\underline{c}\Big) \text{ in prob.,} \tag{2.41}$$

where

$$\Sigma_M^{(u,v)} := \sum\nolimits_{h_1=-2M-m_1+1}^{2M+m_1-1}\sum\nolimits_{h_2=-M-m_2+1}^{M+m_2-1}\operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h},M}), g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big).$$

In order to establish the limiting variance in (2.41), we first recall that $(X^*_{\underline{t},M})$ is strictly stationary as can be seen from (2.18) and the (conditional) i.i.d. property of $(\varepsilon^*_{\underline{t}})$. Therefore, $(\mathbf{Y}^*_{\underline{t},M})$, and consequently $(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}))$, are also strictly stationary processes in $\underline{t}$. Hence, standard calculations yield

$$\begin{aligned} &\operatorname{Var}^*\Big(\frac{1}{n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big)\Big)\\ = \;& \frac{1}{n^2}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{v_1=1}^{n}\sum_{v_2=1}^{n}\operatorname{Cov}^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}), \underline{c}^T g(\mathbf{Y}^*_{\underline{v},M})\Big)\\ = \;& \sum_{h_1=-(n-1)}^{n-1}\sum_{h_2=-(n-1)}^{n-1}\Big(1-\frac{|h_1|}{n}\Big)\Big(1-\frac{|h_2|}{n}\Big)\operatorname{Cov}^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{h},M}), \underline{c}^T g(\mathbf{Y}^*_{\underline{0},M})\Big) \end{aligned}$$

A close inspection of the definition of $\mathbf{Y}^*_{\underline{t},M}$, which depends only on a finite number of random variables $\varepsilon^*_{\underline{t}+\underline{j}}$, together with the i.i.d. property of $(\varepsilon^*_{\underline{t}})$, shows that $g(\mathbf{Y}^*_{\underline{h},M})$ and $g(\mathbf{Y}^*_{\underline{0},M})$ are independent whenever $|h_1| \geq 2M+m_1$ or $|h_2| \geq M+m_2$. Therefore, for all $n \geq \min\{2M+m_1, M+m_2\}$ the last right-hand side equals

$$\begin{aligned} &\sum_{h_1=-2M-m_1+1}^{2M+m_1-1}\sum_{h_2=-M-m_2+1}^{M+m_2-1}\Big(1-\frac{|h_1|}{n}\Big)\Big(1-\frac{|h_2|}{n}\Big)\operatorname{Cov}^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{h},M}), \underline{c}^T g(\mathbf{Y}^*_{\underline{0},M})\Big)\\ = \;& \underline{c}^T\Sigma_M\,\underline{c} + o_P(1), \end{aligned}$$

as can be seen from (2.23). This establishes the correct asymptotic variance in (2.41), and by abbreviating

$$v_n^* := \operatorname{Var}^*\Big(\frac{1}{n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big)\Big)$$

assertion (2.41) follows from Slutsky's Theorem if we can show

$$\frac{1}{n\sqrt{v_n^*}}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big) \xrightarrow{d^*} \mathcal{N}(0,1) \text{ in prob.} \tag{2.42}$$

The strategy is to use a blocking technique. We define sequences of integers $a(n), b(n) \in \mathbb{N}$ with $a(n) \to \infty$, $b(n) \to \infty$ and $b(n)/a(n) \to 0$ as $n \to \infty$. Also, we assume that $a(n)$ and $b(n)$ increase slowly enough such that

$$N(n) := \frac{n}{a(n)+b(n)} \longrightarrow \infty, \quad \text{as } n \to \infty,$$

and assume without loss of generality $N(n) \in \mathbb{N}$ for all n. The idea is to split up the n^2 summands in (2.42) into dominating, square-shaped blocks A_{j_1,j_2} of size

$a(n) \times a(n)$, and negligible remainder terms B_{j_1,j_2} and C_{j_1,j_2}. In the following, we will often abbreviate $a = a(n)$, $b = b(n)$ and $N = N(n)$ in order to simplify the notation. (2.42) can be decomposed as

$$\frac{1}{n\sqrt{v_n^*}} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} \Big(A_{j_1,j_2} + B_{j_1,j_2} + C_{j_1,j_2} \Big) \xrightarrow{d^*} \mathcal{N}(0,1) \text{ in prob.,} \tag{2.43}$$

where

$$\begin{aligned}
A_{j_1,j_2} &:= \sum_{t_1=(j_1-1)(a+b)+1}^{j_1 a+(j_1-1)b} \; \sum_{t_2=(j_2-1)(a+b)+1}^{j_2 a+(j_2-1)b} \Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big), \\
B_{j_1,j_2} &:= \sum_{t_1=j_1 a+(j_1-1)b+1}^{j_1(a+b)} \; \sum_{t_2=(j_2-1)(a+b)+1}^{j_2 a+(j_2-1)b} \Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big), \\
C_{j_1,j_2} &:= \sum_{t_1=(j_1-1)(a+b)+1}^{j_1(a+b)} \; \sum_{t_2=j_2 a+(j_2-1)b+1}^{j_2(a+b)} \Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big).
\end{aligned}$$

We will now establish moment bounds for these three expressions. For the constant h from Assumption 3, define $\delta := 2/(3h+3)$. Writing $A_{1,1} := \sum_{t_1} \eta_{t_1}$ with

$$\eta_{t_1} := \sum_{t_2=1}^{a} \Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) - E^*\Big(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big), \quad t_1 = 1, \ldots, a \,,$$

we get from Theorem 1 in Yokoyama (1980) that $E^*(|\eta_{t_1}|^{2+2\delta}) = a(n)^{1+\delta}\, \mathcal{O}_P(1)$ uniformly in t_1, since η_{t_1} consists of $a(n)$ summands which are centered, $(M+m_2)$-dependent in t_2 and fulfil the required moment assumption $E^*(|\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M})|^{2+3\delta}) = \mathcal{O}_P(1)$ (uniformly) due to (2.22). In other words, we get

$$E^*\Big(\Big|\frac{\eta_{t_1}}{a(n)^{1/2}}\Big|^{2+2\delta}\Big) = \mathcal{O}_P(1),$$

i.e. the $(2M+m_1)$-dependent sequence $(\eta_{t_1}/a(n)^{1/2})_{t_1}$ itself fulfils the conditions of Theorem 1 in Yokoyama (1980) which yields

$$E^*\Big(|A_{1,1}|^{2+\delta}\Big) = a(n)^{1+(\delta/2)} \cdot E^*\Big(\Big| \sum_{t_1=1}^{a} \frac{\eta_{t_1}}{a(n)^{1/2}} \Big|^{2+\delta}\Big) = \Big(a(n) \cdot a(n)\Big)^{1+(\delta/2)} \mathcal{O}_P(1).$$

With analogous calculations we get the bounds

$$\begin{aligned}
E^*\Big(|B_{1,1}|^{2+\delta}\Big) &= \Big(a(n) \cdot b(n)\Big)^{1+(\delta/2)} \mathcal{O}_P(1), \\
E^*\Big(|C_{1,1}|^{2+\delta}\Big) &= \Big((a(n)+b(n)) \cdot b(n)\Big)^{1+(\delta/2)} \mathcal{O}_P(1).
\end{aligned}$$

Note that we can not apply Yokoyama's Theorem directly to $A_{1,1}$, but have to take the intermediate step with η_{t_1} as performed above, because the a^2 many summands in $A_{1,1}$ can *not* be transformed into an M-dependent sequence, regardless of the ordering of the summands.

In the preceeding part of this proof we have shown $v^* = \underline{c}^T \Sigma_M \, \underline{c} + o_P(1)$, and we can assume $\underline{c}^T \Sigma_M \, \underline{c} \neq 0$ (in the case $\underline{c}^T \Sigma_M \, \underline{c} = 0$ the desired assertion (2.41) would follow trivially). Hence, we have $1/v^* = \mathcal{O}_P(1)$. We will now use this assertion, as well as the established moment bounds, to show that B_{j_1,j_2} and C_{j_1,j_2} in (2.43) are asymptotically negligible. The blocks B_{j_1,j_2} are identically distributed and, for n large enough such that $a(n), b(n) > 2M$, independent. Standard calculations yield for any $\varepsilon > 0$:

$$\begin{aligned} P^*\Big\{\Big|\frac{1}{n\sqrt{v_n^*}}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N} B_{j_1,j_2}\Big| > \varepsilon\Big\} &\leq \frac{1}{n^2\varepsilon^2 v^*}\mathrm{Var}^*\Big(\sum_{j_1=1}^{N}\sum_{j_2=1}^{N} B_{j_1,j_2}\Big) \\ &\leq \frac{N^2}{n^2\varepsilon^2 v^*}\Big(E^*\big(|B_{1,1}|^{2+\delta}\big)\Big)^{2/(2+\delta)} \\ &\leq \frac{a(n)\cdot b(n)}{(a(n)+b(n))^2}\,\mathcal{O}_P(1) = o_P(1). \end{aligned} \tag{2.44}$$

The same assertion can be shown for the C_{j_1,j_2}-blocks. Therefore, Slutsky's Theorem implies (2.43) if we can show

$$\frac{\tau_N}{n\sqrt{v_n^*}} \cdot \frac{1}{\tau_N}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N} A_{j_1,j_2} \xrightarrow{d^*} \mathcal{N}(0,1) \ \text{ in prob.}, \tag{2.45}$$

where

$$\tau_N := \Big(\sum_{j_1=1}^{N}\sum_{j_2=1}^{N} \mathrm{Var}^*(A_{j_1,j_2})\Big)^{1/2}.$$

Observe that, for n large enough such that $b(n) > 2M$, the A_{j_1,j_2}-blocks are independent random variables, and in the following we will only consider those n. Per definition, we can decompose

$$\begin{aligned} \frac{\tau_N^2}{n^2} &= v_n^* + \mathrm{Var}^*\Big(\frac{1}{n}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}(B_{j_1,j_2}+C_{j_1,j_2})\Big) \\ &\quad -2\,\mathrm{Cov}^*\Big(\frac{1}{n}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}(A_{j_1,j_2}+B_{j_1,j_2}+C_{j_1,j_2}), \frac{1}{n}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N}(B_{j_1,j_2}+C_{j_1,j_2})\Big) \\ &= v_n^* + o_P(1), \end{aligned} \tag{2.46}$$

which follows from $n^{-2}\operatorname{Var}^*(\sum_{j_1}\sum_{j_2}(B_{j_1,j_2}+C_{j_1,j_2})) = o_P(1)$, cf. the calculations leading up to (2.44). Therefore, the first factor on the left-hand side of (2.45) converges to 1 in probability.

We will apply Lindeberg's central limit theorem to the second factor in (2.45), recalling that the A_{j_1,j_2}-blocks are i.i.d. random variables for n large enough. Using $N = (n/a(n))\cdot\mathcal{O}(1)$ and the moment condition for $A_{1,1}$ established earlier, as well as the fact that (2.46) implies $n/\tau_N = \mathcal{O}_P(1)$, we can check the Lyapunov condition:

$$\begin{aligned} &\frac{1}{\tau_N^{2+\delta}}\sum_{j_1=1}^{N}\sum_{j_2=1}^{N} E^*\Big(|A_{j_1,j_2}|^{2+\delta}\Big) = \frac{N^2}{\tau_N^{2+\delta}}\, a(n)^{2+\delta}\,\mathcal{O}_P(1)\\ =\ &\left(\frac{n}{\tau_N}\right)^2\left(\frac{a(n)}{n}\cdot\frac{n}{\tau_N}\right)^{\delta}\mathcal{O}_P(1) = \left(\frac{n}{\tau_N}\right)^{2+\delta}\left(\frac{1}{N(n)}\right)^{\delta}\mathcal{O}_P(1) = o_P(1). \end{aligned}$$

This yields (2.45) and, consequently, (2.41).

We will invoke Proposition 6.3.9 of Brockwell and Davis (1991) to show that (2.41) implies (2.40). The next step is to prove $\Sigma_M^{(u,v)} \to \Sigma^{(u,v)}$, as $M\to\infty$. It holds

$$\begin{aligned} \Sigma_M^{(u,v)} =& \sum_{h_1=-2M-m_1+1}^{2M+m_1-1}\ \sum_{h_2=-M-m_2+1}^{M+m_2-1} \operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\\ &+ \sum_{h_1=-2M-m_1+1}^{2M+m_1-1}\ \sum_{h_2=-M-m_2+1}^{M+m_2-1}\Big(\operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h},M}), g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big) - \operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\Big). \end{aligned}$$

The first summand on the right-hand side converges to $\Sigma^{(u,v)}$, as $M\to\infty$, due to (2.24). As for the second summand, we have

$$\begin{aligned} &\sum_{h_1=-2M-m_1+1}^{2M+m_1-1}\ \sum_{h_2=-M-m_2+1}^{M+m_2-1}\Big|\operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h},M}), g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big) - \operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\Big|\\ \le& \sum_{h_1}\sum_{h_2}\Big|\operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h},M}) - g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big) + \operatorname{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0},M}) - g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\Big|\\ \le& \sum_{h_1}\sum_{h_2}\Big(\Big\|g_u(\widetilde{\mathbf{Y}}_{\underline{h},M}) - g_u(\widetilde{\mathbf{Y}}_{\underline{h}})\Big\|_2\Big\|g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big\|_2 + \Big\|g_u(\widetilde{\mathbf{Y}}_{\underline{h}})\Big\|_2\Big\|g_v(\widetilde{\mathbf{Y}}_{\underline{0},M}) - g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big\|_2\Big)\\ \le& C\cdot \sum_{h_1=-2M-m_1+1}^{2M+m_1-1}\ \sum_{h_2=-M-m_2+1}^{M+m_2-1}\ \sum_{\underline{k}\in\Theta\setminus\Theta(M)} |b_{\underline{k}}|, \end{aligned}\tag{2.47}$$

for some generic constant $C<\infty$ (which, from now on, may change from line to line). The latter inequality can be derived from Lemma 2.14 (with $W=\Theta(M)\cup\{\underline{0}\}$), using the fact that

$$\Big\|g_u(\widetilde{\mathbf{Y}}_{\underline{h}})\Big\|_2 \le \Big\|g_u(\widetilde{\mathbf{Y}}_{\underline{h}}) - g_u(\widetilde{\mathbf{Y}}_{\underline{h}}^{(\emptyset)})\Big\|_2 + \Big\|g_u(\underline{0})\Big\|_2 \le 1 + \sum_{\underline{k}\in\Theta}|b_{\underline{k}}| + \Big|g_u(\underline{0})\Big| < \infty$$

follows also from Lemma 2.14 (with $W=\emptyset$), the same being true for $\|g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})\|_2$. For (2.47) it holds

$$\begin{aligned}
& C\cdot\sum_{h_1=-2M-m_1+1}^{2M+m_1-1}\ \sum_{h_2=-M-m_2+1}^{M+m_2-1}\ \sum_{\underline{k}\in\Theta\setminus\Theta(M)}|b_{\underline{k}}| \\
\leq\ & C\cdot\sum_{\underline{k}\in\Theta\setminus\Theta(M)} M^2\,|b_{\underline{k}}| \\
\leq\ & C\cdot\left(\sum_{k_1=M+1}^{\infty}\sum_{k_2=0}^{M}|k_1|^2\,|b_{\underline{k}}|+\sum_{k_1=-\infty}^{-M-1}\sum_{k_2=0}^{M}|k_1|^2\,|b_{\underline{k}}|+\sum_{k_1=-\infty}^{\infty}\sum_{k_2=M+1}^{\infty}|k_2|^2\,|b_{\underline{k}}|\right) \\
\leq\ & C\cdot\sum_{\underline{k}\in\Theta\setminus\Theta(M)}(1+|\underline{k}|_\infty)^2\,|b_{\underline{k}}|,
\end{aligned}$$

which converges to zero, as $M\to\infty$, due to (2.6) and the assumption $r=4$. Hence, we have shown that $\Sigma_M^{(u,v)}\to\Sigma^{(u,v)}$, as $M\to\infty$.

Now we apply Proposition 6.3.9 of Brockwell and Davis (1991) to the bootstrap quantities from (2.41), i.e. we show that it holds

$$\lim_{M\to\infty}\limsup_{n\to\infty}P^*\left\{\left|\frac{1}{n}\sum_{t_1,t_2=1}^{n}\underline{c}^T\Big(g(\mathbf{Y}^*_{\underline{t}})-g(\mathbf{Y}^*_{\underline{t},M})-E^*\Big(g(\mathbf{Y}^*_{\underline{t}})-g(\mathbf{Y}^*_{\underline{t},M})\Big)\Big)\right|>\delta\right\}=0 \quad \text{in } P\text{-probability,} \tag{2.48}$$

for any $\delta>0$. Then (2.40) follows from said Proposition 6.3.9, using the Cramér-Wold device. For condition (2.48) to hold, it is sufficient to show

$$\mathrm{Var}^*\Big(\frac{1}{n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\Big(\underline{c}^Tg(\mathbf{Y}^*_{\underline{t}})-\underline{c}^Tg(\mathbf{Y}^*_{\underline{t},M})\Big)\Big)\leq\frac{1}{M^2}\,\mathcal{O}_P(1). \tag{2.49}$$

We abbreviate $Z^*_{\underline{t},M}:=\underline{c}^Tg(\mathbf{Y}^*_{\underline{t}})-\underline{c}^Tg(\mathbf{Y}^*_{\underline{t},M})$ and observe that $(Z^*_{\underline{t},M})$ is a stationary spatial process. Standard calculations yield

$$\begin{aligned}
\mathrm{Var}^*\Big(\frac{1}{n}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}Z^*_{\underline{t},M}\Big) &= \sum_{h_1=-(n-1)}^{n-1}\ \sum_{h_2=-(n-1)}^{n-1}\frac{(n-|h_1|)(n-|h_2|)}{n^2}\,\mathrm{Cov}^*(Z^*_{\underline{0},M},Z^*_{\underline{h},M}) \\
&\leq \sum_{h_1=-\infty}^{\infty}\sum_{h_2=-\infty}^{\infty}\Big|\mathrm{Cov}^*(Z^*_{\underline{0},M},Z^*_{\underline{h},M})\Big| \\
&\leq 2\sum_{h_1=-\infty}^{\infty}\sum_{h_2=0}^{\infty}\Big|\mathrm{Cov}^*(Z^*_{\underline{0},M},Z^*_{\underline{h},M})\Big|,
\end{aligned}$$

noting that $\mathrm{Cov}^*(Z^*_{\underline{0},M},Z^*_{\underline{h},M})=\mathrm{Cov}^*(Z^*_{\underline{0},M},Z^*_{-\underline{h},M})$. Hence, to obtain (2.49), it suffices to show

$$\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\Big|\mathrm{Cov}^*(Z^*_{\underline{0},M},Z^*_{\underline{h},M})\Big|\leq\frac{1}{M^2}\,\mathcal{O}_P(1), \tag{2.50}$$

since the remaining part $\sum_{h_1=-\infty}^{-1}\sum_{h_2=0}^{\infty}|\mathrm{Cov}^*(Z^*_{\underline{0},M},Z^*_{\underline{h},M})|$ can be treated analogously. In order to show (2.50) we will make use of three different truncated versions of $Z^*_{\underline{t},M}$, which we will denote by $Z^{*[1]}_{\underline{t},M},\ldots,Z^{*[3]}_{\underline{t},M}$. The truncation points will depend on h_1 and h_2, the indices showing up in (2.50), which will be suppressed in the notation. Each of the truncated versions is generated in a natural way from truncated versions of $X^*_{\underline{t}}$ and $X^*_{\underline{t},M}$. To be precise, we set

$$Z^{*[j]}_{\underline{t},M} := \underline{c}^T g(\mathbf{Y}^{*[j]}_{\underline{t}}) - \underline{c}^T g(\mathbf{Y}^{*[j]}_{\underline{t},M}), \quad j=1,2,3\,,$$

where

$$\mathbf{Y}^{*[j]}_{\underline{t}} := (X^{*[j]}_{\underline{t}+\underline{s}(1)},\ldots,X^{*[j]}_{\underline{t}+\underline{s}(m_1m_2)})^T, \quad \mathbf{Y}^{*[j]}_{\underline{t},M} := (X^{*[j]}_{\underline{t}+\underline{s}(1),M},\ldots,X^{*[j]}_{\underline{t}+\underline{s}(m_1m_2),M})^T$$

and, setting $\widehat{b}_{(0,0)}(p) := 1$ and $\widehat{b}_{(k_1,0)}(p) := 0$ for $k_1<0$,

$$\begin{aligned}
X^{*[1]}_{\underline{t}} &:= \sum_{k_1=-\infty}^{\infty}\sum_{k_2=0}^{h_2-m_2}\widehat{b}_{\underline{k}}(p)\,\varepsilon^*_{\underline{t}-\underline{k}}\cdot\mathbb{1}_{\{h-m_2\geq 0\}} + \sum_{k_1=-\infty}^{\lfloor h_1/2\rfloor}\sum_{k_2=(h_2-m_2+1)\vee 0}^{\infty}\widehat{b}_{\underline{k}}(p)\,\varepsilon^*_{\underline{t}-\underline{k}},\\
X^{*[2]}_{\underline{t}} &:= \sum_{k_1=-\lfloor h_1/2\rfloor+m_1}^{\infty}\sum_{k_2=0}^{\infty}\widehat{b}_{\underline{k}}(p)\,\varepsilon^*_{\underline{t}-\underline{k}},\\
X^{*[3]}_{\underline{t}} &:= \sum_{k_1=-\infty}^{\infty}\sum_{k_2=0}^{h_2-m_2}\widehat{b}_{\underline{k}}(p)\,\varepsilon^*_{\underline{t}-\underline{k}}\cdot\mathbb{1}_{\{h-m_2\geq 0\}}.
\end{aligned}$$

The versions $X^{*[j]}_{\underline{t},M}$, $j=1,2,3$, can be obtained from the corresponding definitions of $X^{*[j]}_{\underline{t}}$, by replacing each $\widehat{b}_{\underline{k}}(p)$ with $\widehat{b}_{\underline{k}}(p)\cdot\mathbb{1}_{\{\underline{k}\in\Theta(M)\}}$.
With these definitions we can now split up the expression in (2.50) as

$$\begin{aligned}
&\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left|\mathrm{Cov}^*(Z^*_{\underline{0},M},Z^*_{\underline{h},M})\right|\\
\leq\;&\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left|\mathrm{Cov}^*(Z^*_{\underline{0},M},Z^*_{\underline{h},M}-Z^{*[1]}_{\underline{h},M})\right| + \sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left|\mathrm{Cov}^*(Z^{*[2]}_{\underline{0},M},Z^{*[1]}_{\underline{h},M})\right|\\
&+\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left|\mathrm{Cov}^*(Z^*_{\underline{0},M}-Z^{*[2]}_{\underline{0},M},Z^{*[3]}_{\underline{h},M})\right|\\
&+\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left|\mathrm{Cov}^*(Z^*_{\underline{0},M}-Z^{*[2]}_{\underline{0},M},Z^{*[1]}_{\underline{h},M}-Z^{*[3]}_{\underline{h},M})\right|\\
=:\;& I+II+III+IV.
\end{aligned}$$

A close inspection of the definition of the different truncated versions $Z^{*[j]}_{\underline{0},M}$ and $Z^{*[j]}_{\underline{h},M}$ shows that $Z^{*[2]}_{\underline{0},M}$ and $Z^{*[1]}_{\underline{h},M}$ are independent random variables because they depend on disjoint sets of variables $\varepsilon^*_{\underline{t}}$ (this is why the truncated versions are defined as they are), and the $(\varepsilon^*_{\underline{t}})$ are i.i.d.. With the same argument, $Z^{*[3]}_{\underline{h},M}$ is independent of both $Z^*_{\underline{0},M}$ and $Z^{*[2]}_{\underline{0},M}$. Thus, the expressions II and III are identical zero. We can therefore prove (2.50) by showing

$$I \leq \frac{1}{M^2}\, \mathcal{O}_P(1), \quad IV \leq \frac{1}{M^2}\, \mathcal{O}_P(1). \tag{2.51}$$

Using the notation $\|\cdot\|_{*2}$ introduced in Lemma 2.14, we have for I

$$\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left|\mathrm{Cov}^*(Z^*_{\underline{0},M}, Z^*_{\underline{h},M} - Z^{*[1]}_{\underline{h},M})\right| \leq \left\|Z^*_{\underline{0},M}\right\|_{*2}\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left\|Z^*_{\underline{h},M} - Z^{*[1]}_{\underline{h},M}\right\|_{*2},$$

and

$$\begin{aligned}
&\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left\|Z^*_{\underline{h},M} - Z^{*[1]}_{\underline{h},M}\right\|_{*2} \\
\leq\;& \sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\left(\left\|\underline{c}^T g(\mathbf{Y}^*_{\underline{h}}) - \underline{c}^T g(\mathbf{Y}^{*[1]}_{\underline{h}})\right\|_{*2} + \left\|\underline{c}^T g(\mathbf{Y}^*_{\underline{h},M}) - \underline{c}^T g(\mathbf{Y}^{*[1]}_{\underline{h},M})\right\|_{*2}\right) \\
\leq\;& \sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\sum_{u=1}^{k}|c_u|\left(\left\|g_u(\mathbf{Y}^*_{\underline{h}}) - g_u(\mathbf{Y}^{*[1]}_{\underline{h}})\right\|_{*2} + \left\|g_u(\mathbf{Y}^*_{\underline{h},M}) - g_u(\mathbf{Y}^{*[1]}_{\underline{h},M})\right\|_{*2}\right) \\
\leq\;& \sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\sum_{u=1}^{k}|c_u|\left(\mathcal{O}_P(1)\cdot\sum_{k_1=\lfloor h_1/2\rfloor+1}^{\infty}\;\sum_{k_2=(h_2-m_2+1)\vee 0}^{\infty}\left|\widehat{b}_{\underline{k}}(p)\right|\right)
\end{aligned} \tag{2.52}$$

where the $\mathcal{O}_P(1)$-expression on the right-hand side does not depend on $\underline{h}$, u or M. The latter inequality follows directly from Lemma 2.14. The last right-hand side can be bounded by

$$\begin{aligned}
&\mathcal{O}_P(1)\cdot\sum_{h_1=0}^{\infty}\sum_{h_2=0}^{\infty}\sum_{k_1=\lfloor h_1/2\rfloor+1}^{\infty}\;\sum_{k_2=(h_2-m_2+1)\vee 0}^{\infty}\left|\widehat{b}_{\underline{k}}(p)\right| \\
\leq\;& \mathcal{O}_P(1)\cdot m_2\sum_{h_1=0}^{\infty}\sum_{h_2=m_2-1}^{\infty}\sum_{k_1=\lfloor h_1/2\rfloor+1}^{\infty}\;\sum_{k_2=h_2-m_2+1}^{\infty}\left|\widehat{b}_{\underline{k}}(p)\right| \\
\leq\;& \mathcal{O}_P(1)\cdot m_2\sum_{h_1=0}^{\infty}\sum_{k_1=\lfloor h_1/2\rfloor+1}^{\infty}\sum_{k_2=0}^{\infty}(k_2+1)\left|\widehat{b}_{\underline{k}}(p)\right| \\
\leq\;& \mathcal{O}_P(1)\cdot m_2\sum_{k_1=1}^{\infty}\sum_{k_2=0}^{\infty}2k_1\,(k_2+1)\left|\widehat{b}_{\underline{k}}(p)\right| \\
\leq\;& \mathcal{O}_P(1)\cdot 2m_2\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2\left|\widehat{b}_{\underline{k}}(p)\right| = \mathcal{O}_P(1),
\end{aligned}$$

using (2.19). Thus, we have

$$I \leq \left\|Z^*_{\underline{0},M}\right\|_{*2}\cdot\mathcal{O}_P(1).$$

We also get from Lemma 2.14

$$
\begin{aligned}
M^2 \cdot \left\| Z^*_{\underline{0},M} \right\|_{*2} &\leq M^2 \cdot \sum_{u=1}^{k} |c_u| \left\| g_u(\mathbf{Y}^*_{\underline{0}}) - g_u(\mathbf{Y}^*_{\underline{0},M}) \right\|_{*2} \\
&\leq \mathcal{O}_P(1) \cdot M^2 \sum_{\underline{k} \in \Theta \setminus \Theta(M)} \left| \widehat{b}_{\underline{k}}(p) \right| \\
&\leq \mathcal{O}_P(1) \cdot \sum_{\underline{k} \in \Theta} (1 + |\underline{k}|_\infty)^2 \left| \widehat{b}_{\underline{k}}(p) \right| = \mathcal{O}_P(1), \qquad (2.53)
\end{aligned}
$$

uniformly for all $M \in \mathbb{N}$, due to (2.19) and $M < |\underline{k}|_\infty$ for all $\underline{k} \in \Theta \setminus \Theta(M)$. It follows

$$
I \leq \frac{1}{M^2} \mathcal{O}_P(1).
$$

Turning to expression IV, we can decompose

$$
\begin{aligned}
IV &\leq \sum_{h_1=0}^{M} \sum_{h_2=0}^{M} \left| \mathrm{Cov}^*(Z^*_{\underline{0},M} - Z^{*[2]}_{\underline{0},M}, Z^{*[1]}_{\underline{h},M} - Z^{*[3]}_{\underline{h},M}) \right| \\
&\quad + \sum_{h_1=M+1}^{\infty} \sum_{h_2=M+1}^{\infty} \left| \dots \right| + \sum_{h_1=0}^{M} \sum_{h_2=M+1}^{\infty} \left| \dots \right| + \sum_{h_1=M+1}^{\infty} \sum_{h_2=0}^{M} \left| \dots \right| \\
&=: A + B + C + D.
\end{aligned}
$$

With the same techniques as in (2.53) we get

$$
\begin{aligned}
M^2 \cdot A &\leq M^2 \cdot \sum_{h_1=0}^{M} \sum_{h_2=0}^{M} \left(\left\| Z^*_{\underline{0},M} \right\|_{*2} + \left\| Z^{*[2]}_{\underline{0},M} \right\|_{*2} \right) \left(\left\| Z^{*[1]}_{\underline{h},M} \right\|_{*2} + \left\| Z^{*[3]}_{\underline{h},M} \right\|_{*2} \right) \\
&\leq \mathcal{O}_P(1) \cdot M^2 \sum_{h_1=0}^{M} \sum_{h_2=0}^{M} \left(2 \sum_{\underline{k} \in \Theta \setminus \Theta(M)} \left| \widehat{b}_{\underline{k}}(p) \right| \right)^2 \\
&\leq \mathcal{O}_P(1) \cdot \left(\sum_{\underline{k} \in \Theta \setminus \Theta(M)} (M+1)^2 \left| \widehat{b}_{\underline{k}}(p) \right| \right)^2 \\
&\leq \mathcal{O}_P(1) \cdot \left(\sum_{\underline{k} \in \Theta} (1 + |\underline{k}|_\infty)^2 \left| \widehat{b}_{\underline{k}}(p) \right| \right)^2 = \mathcal{O}_P(1) \qquad (2.54)
\end{aligned}
$$

uniformly for all $M \in \mathbb{N}$. Since we are interested in an asymptotic result for $M \to \infty$ in (2.48), we can, from now on, consider only those M large enough such that $-\lfloor (M+1)/2 \rfloor + m_1 - 1 < 0$ and $M - m_2 + 2 \geq 0$. With the same calculation as for $\|Z^*_{\underline{h},M} - Z^{*[1]}_{\underline{h},M}\|_{*2}$ in (2.52), we can derive

$$
\left\| Z^*_{\underline{0},M} - Z^{*[2]}_{\underline{0},M} \right\|_{*2} \leq \mathcal{O}_P(1) \cdot \sum_{k_1=-\infty}^{-\lfloor h_1/2 \rfloor + m_1 - 1} \sum_{k_2=0}^{\infty} \left| \widehat{b}_{\underline{k}}(p) \right|,
$$

$$
\begin{aligned}
\left\| Z^{*[1]}_{\underline{h},M} - Z^{*[3]}_{\underline{h},M} \right\|_{*2} &\leq \mathcal{O}_P(1) \cdot \sum_{k_1=-\infty}^{\lfloor h_1/2 \rfloor} \sum_{k_2=(h_2-m_2+1)\vee 0}^{\infty} \left| \widehat{b}_{\underline{k}}(p) \right| \\
&\leq \mathcal{O}_P(1) \cdot \sum_{k_1=-\infty}^{\infty} \sum_{k_2=(h_2-m_2+1)\vee 0}^{\infty} \left| \widehat{b}_{\underline{k}}(p) \right|.
\end{aligned}
$$

With these bounds and for all M large enough as defined before, we get for B:

$$
\begin{aligned}
& M^2 \cdot B \\
\leq\ & M^2 \cdot \sum_{h_1=M+1}^{\infty} \sum_{h_2=M+1}^{\infty} \left\| Z^{*}_{\underline{0},M} - Z^{*[2]}_{\underline{0},M} \right\|_{*2} \left\| Z^{*[1]}_{\underline{h},M} - Z^{*[3]}_{\underline{h},M} \right\|_{*2} \\
\leq\ & \mathcal{O}_P(1) \cdot M^2 \sum_{h_1=M+1}^{\infty} \sum_{k_1=-\infty}^{-\lfloor h_1/2 \rfloor + m_1 - 1} \sum_{k_2=0}^{\infty} \left| \widehat{b}_{\underline{k}}(p) \right| \cdot \sum_{h_2=M+1}^{\infty} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=h_2-m_2+1}^{\infty} \left| \widehat{b}_{\underline{k}}(p) \right| \\
=\ & \mathcal{O}_P(1) \cdot B_1 \cdot B_2,
\end{aligned}
$$

where

$$
\begin{aligned}
B_1 &:= M \sum_{h_1=M+1}^{\infty} \sum_{k_1=-\infty}^{-\lfloor h_1/2 \rfloor + m_1 - 1} \sum_{k_2=0}^{\infty} \left| \widehat{b}_{\underline{k}}(p) \right| \\
&= M \sum_{k_1=-\infty}^{-\lfloor (M+1)/2 \rfloor + m_1 - 1} \sum_{k_2=0}^{\infty} 2 \left| k_1 - \Big(-\lfloor (M+1)/2 \rfloor + m_1 \Big) \right| \left| \widehat{b}_{\underline{k}}(p) \right| \\
&\leq 2 \sum_{k_1=-\infty}^{-\lfloor (M+1)/2 \rfloor + m_1 - 1} \sum_{k_2=0}^{\infty} M \left| k_1 \right| \left| \widehat{b}_{\underline{k}}(p) \right| \\
&\leq 8 m_1 \sum_{\underline{k} \in \Theta} (1 + |\underline{k}|_\infty)^2 \left| \widehat{b}_{\underline{k}}(p) \right| = \mathcal{O}_P(1),
\end{aligned}
\tag{2.55}
$$

and

$$
\begin{aligned}
B_2 &:= M \sum_{h_2=M+1}^{\infty} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=h_2-m_2+1}^{\infty} \left| \widehat{b}_{\underline{k}}(p) \right| \\
&= M \sum_{k_1=-\infty}^{\infty} \sum_{k_2=M-m_2+2}^{\infty} \left| k_2 - (M - m_2 + 1) \right| \left| \widehat{b}_{\underline{k}}(p) \right| \\
&\leq \sum_{k_1=-\infty}^{\infty} \sum_{k_2=M-m_2+2}^{\infty} M \left| k_2 \right| \left| \widehat{b}_{\underline{k}}(p) \right| \\
&\leq (m_2 + 1) \sum_{\underline{k} \in \Theta} (1 + |\underline{k}|_\infty)^2 \left| \widehat{b}_{\underline{k}}(p) \right| = \mathcal{O}_P(1),
\end{aligned}
\tag{2.56}
$$

uniformly for all M large enough. The latter inequalities in (2.55) and (2.56) hold because

$$
M \left| k_2 \right| \leq (k_2 + m_2 - 2)\, k_2 \leq (m_2 + 1)(k_2 + 1)\, k_2 \leq (m_2 + 1)\, (1 + |\underline{k}|_\infty)^2
$$

for all $k_2 \geq M - m_2 + 2$, and, with similar calculations,

$$M\,|k_1| \leq 2m_1\,(2|k_1|+1)\,|k_1| \leq 4m_1\,(1+|k_1|)\,|k_1| \leq 4m_1\,(1+|\underline{k}|_\infty)^2$$

for all $k_1 \leq -\lfloor (M+1)/2 \rfloor + m_1 - 1$. Altogether, this yields $B \leq (1/M^2)\,\mathcal{O}_P(1)$, and, with exactly the same arguments as for A and B, we can also show $C \leq (1/M^2)\,\mathcal{O}_P(1)$ and $D \leq (1/M^2)\,\mathcal{O}_P(1)$. This implies (2.51) and, therefore, (2.50) and (2.49). Hence, the proof of (2.38) is completed.

Using analogous arguments as for the bootstrap quantities in (2.38), one can show (2.39) for the non-bootstrap quantities. Since

$$\left\| \underline{\theta}^* - \widetilde{\underline{\theta}} \right\| := \sum_{v=1}^{k} \left| E^*\Big(g_v(\mathbf{Y}^*_{\underline{t}})\Big) - E\Big(g_v(\widetilde{\mathbf{Y}}_{\underline{t}})\Big) \right| = o_P(1) \tag{2.57}$$

follows with the same arguments as in the proof of (2.23) (by simply replacing covariances with expectations), we can incorporate the outer function f from the definition of $\widetilde{T}_n$ and T^*_n, cf. Assumption 3, with the delta method. It follows from (2.38), (2.39) and (2.57) that $(\bar{n}_1\bar{n}_2)^{1/2}(T^*_n - f(\underline{\theta}^*))$ and $(\bar{n}_1\bar{n}_2)^{1/2}(\widetilde{T}_n - f(\widetilde{\underline{\theta}}))$ have identical limiting (normal) distributions. Therefore, since $(\bar{n}_1\bar{n}_2)^{1/2}$ is asymptotically equivalent to n, we have

$$\sup_{x\in\mathbb{R}} \left| P^*\Big\{ n(T^*_n - f(\underline{\theta}^*)) \leq x \Big\} - P\Big\{ n(\widetilde{T}_n - f(\widetilde{\underline{\theta}})) \leq x \Big\} \right| = o_P(1),$$

which completes the proof. □

2.7 Proofs of the auxiliary results

Proof of Lemma 2.1:

Finding one-sided AR- and MA-representations as in (2.5) is closely related to finding a spectral factorization $f(\underline{\lambda}) = |B'(\underline{\lambda})|^2$ of the spectral density, where B' is a complex-valued function with one-sided Fourier series in the sense of the half-plane Θ, i.e.

$$B'(\underline{\lambda}) = \sum_{\underline{k}\in\Theta\cup\{\underline{0}\}} \widetilde{b}_{\underline{k}}\, e^{-i\langle \underline{k},\underline{\lambda}\rangle}. \tag{2.58}$$

Under Assumption 1 the spectral density f is equal to its absolutely convergent Fourier series $f(\underline{\lambda}) = \sum_{\underline{k}\in\mathbb{Z}^2} (\gamma(\underline{k})/4\pi^2)\, e^{i\langle \underline{k},\underline{\lambda}\rangle}$. Lemma 2.3 shows that $\log f \in C_{r-1}$

and, in particular, $\log f$ is equal to its absolutely convergent Fourier series,

$$\log f(\underline{\lambda}) = \sum_{\underline{k}\in\mathbb{Z}^2} d_{\underline{k}}\, e^{i\langle \underline{k},\underline{\lambda}\rangle},$$

say. Whittle (1954) showed that the spectral factorization of f can be obtained from the Fourier series of $\log f$ by letting

$$B_0(\underline{z}) := \exp\Big(L(\underline{z})\Big), \quad L(\underline{z}) := \frac{d_{\underline{0}}}{2} + \sum_{\underline{k}\in\Theta} d_{\underline{k}}\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S, \tag{2.59}$$

where $S = \{\underline{z} \in \mathbb{C}^2,\ |z_1| = 1,\ |z_2| \leq 1\}$. Identifying $B'(\underline{\lambda}) := B_0(e^{-i\lambda_1}, e^{-i\lambda_2})$ gives the spectral factorization $f(\underline{\lambda}) = |B'(\underline{\lambda})|^2$, since one can easily verify

$$B_0(e^{-i\lambda_1}, e^{-i\lambda_2}) \cdot \overline{B_0(e^{-i\lambda_1}, e^{-i\lambda_2})} = f(\underline{\lambda})$$

(note that $\log f(-\underline{\lambda}) = \log f(\underline{\lambda})$ implies $d_{\underline{k}} = d_{-\underline{k}} \in \mathbb{R}$ for all $\underline{k} \in \mathbb{Z}^2$). Through straightforward multiplication and grouping of like summands one can easily verify that each power $L(\underline{z})^j$, $j \in \mathbb{N}_0$, has a series representation with respect to the upper half-plane $\Theta \cup \{\underline{0}\}$, only, i.e.

$$L(\underline{z})^j := \frac{d_{\underline{0}}(j)}{2} + \sum_{\underline{k}\in\Theta} d_{\underline{k}}(j)\, z_1^{k_1} z_2^{k_2} \quad \forall\, \underline{z} \in S.$$

Using this, and expanding $B_0(\underline{z})$ in (2.59) via $\exp(L(\underline{z})) = \sum_{j=0}^{\infty} L(\underline{z})^j/j!$ yields

$$B_0(\underline{z}) = \sum_{\underline{k}\in\Theta\cup\{\underline{0}\}} \widetilde{b}_{\underline{k}}\, z_1^{k_1} z_2^{k_2}$$

for suitable coefficients $\widetilde{b}_{\underline{k}}$. $B'(\underline{\lambda}) = B_0(e^{-i\lambda_1}, e^{-i\lambda_2})$ then gives the desired form (2.58) of the Fourier series of $B'(\underline{\lambda})$. The series $B_0(\underline{z})$, and therefore also the series in (2.58), converge absolutely because the Fourier series of $\log f$ converges absolutely and the power series of the exponential function converges absolutely in $\mathbb{C}$. Furthermore, it holds $|z_1^{k_1}| \leq 1$, $|z_2^{k_2}| \leq 1$ for all $\underline{z} \in S$ and all $\underline{k} \in \Theta$ (note that Θ contains only vectors $\underline{k}$ with $k_2 \geq 0$). From (2.59) it is obvious that $B_0(\underline{z}) \neq 0$ for all $\underline{z}$ with $|z_1| = 1$, $|z_2| \leq 1$ and we can define $A_0(\underline{z}) := 1/B_0(\underline{z})$ on this region which, analogously to B_0, has a one-sided series representation with absolutely summable coefficients $(\widetilde{a}_{\underline{k}})_{\underline{k}\in\Theta\cup\{\underline{0}\}}$, i.e.

$$A_0(\underline{z}) = \frac{1}{B_0(\underline{z})} = \exp\left(\frac{-d_{\underline{0}}}{2} + \sum_{\underline{k}\in\Theta} (-d_{\underline{k}})\, z_1^{k_1} z_2^{k_2}\right) = \sum_{\underline{k}\in\Theta\cup\{\underline{0}\}} \widetilde{a}_{\underline{k}}\, z_1^{k_1} z_2^{k_2}. \tag{2.60}$$

From the definitions of A_0 and B_0 it follows immediately $\widetilde{a}_{\underline{0}} = \exp(-d_{\underline{0}}/2) \neq 0$ and $\widetilde{b}_{\underline{0}} = \exp(d_{\underline{0}}/2) \neq 0$ and we get the standardized versions

$$A(\underline{z}) := \frac{A_0(\underline{z})}{\widetilde{a}_{\underline{0}}} = 1 - \sum_{\underline{k} \in \Theta} a_{\underline{k}}\, z_1^{k_1} z_2^{k_2}, \quad B(\underline{z}) := \frac{B_0(\underline{z})}{\widetilde{b}_{\underline{0}}} = 1 + \sum_{\underline{k} \in \Theta} b_{\underline{k}}\, z_1^{k_1} z_2^{k_2}, \quad (2.61)$$

where $a_{\underline{k}} := -\widetilde{a}_{\underline{k}}/\widetilde{a}_{\underline{0}}$ and $b_{\underline{k}} := \widetilde{b}_{\underline{k}}/\widetilde{b}_{\underline{0}}$ for all $\underline{k} \in \Theta$. (2.61) yields exactly the z-transforms defined in (2.10). We now consider the functions A' and L' on $(-\pi, \pi]^2$ which are, just as B_0 and B', defined via $A'(\lambda_1, \lambda_2) := A_0(e^{-i\lambda_1}, e^{-i\lambda_2})$, $L'(\lambda_1, \lambda_2) := L(e^{-i\lambda_1}, e^{-i\lambda_2})$. Per definition, it holds $B'(\underline{\lambda}) = \exp(L'(\underline{\lambda}))$ and $A'(\underline{\lambda}) = \exp(-L'(\underline{\lambda}))$. Using the submultiplicative C_{r-1}-norm defined in Lemma 2.3, as well as the fact that $\log f \in C_{r-1}$ implies $L' \in C_{r-1}$, we can infer

$$\|B'\|_{r-1} = \Big\| \sum_{j=0}^{\infty} \frac{1}{j!} (L')^j \Big\|_{r-1} \leq \sum_{j=0}^{\infty} \frac{1}{j!} \|L'\|_{r-1}^j = \exp(\|L'\|_{r-1}) < \infty.$$

Analogously, the same argument delivers $A' \in C_{r-1}$ which yields (2.6).
We can now define the process $(\varepsilon_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ via

$$\varepsilon_{\underline{t}} := X_{\underline{t}} - \sum_{\underline{k} \in \Theta} a_{\underline{k}}\, X_{\underline{t}-\underline{k}}$$

which is obviously an L^2-convergent series with spectral density

$$f_\varepsilon(\underline{\lambda}) = |\widetilde{a}_{\underline{0}}^{-1}\, A'(\underline{\lambda})|^2 \cdot f(\underline{\lambda}) = \widetilde{a}_{\underline{0}}^{-2}, \quad \forall\, \underline{\lambda} \in (-\pi, \pi]^2,$$

since $f(\underline{\lambda}) = |B'(\underline{\lambda})|^2 = 1/|A'(\underline{\lambda})|^2$. Hence, $(\varepsilon_{\underline{t}})$ is uncorrelated white noise. Furthermore, the backshift operators of $\varepsilon_{\underline{t}}$ and $X_{\underline{t}}$ coincide and $1/(\widetilde{a}_{\underline{0}}^{-1}\, A'(\underline{\lambda})) = \widetilde{b}_{\underline{0}}^{-1}\, B'(\underline{\lambda})$ implies

$$X_{\underline{t}} = \sum_{\underline{k} \in \Theta} b_{\underline{k}}\, \varepsilon_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}}.$$

It remains to show that $(\varepsilon_{\underline{t}})$ is the innovation process, i.e. that $\sum_{\underline{k} \in \Theta} a_{\underline{k}}\, X_{\underline{t}-\underline{k}}$ is the L^2-projection of $X_{\underline{t}}$ onto $H_{\underline{t}}(X) := \overline{sp}\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta\}$. Let $\underline{j} \in \Theta$ be arbitrary. We show that $X_{\underline{t}} - \sum_{\underline{k} \in \Theta} a_{\underline{k}}\, X_{\underline{t}-\underline{k}}$ is orthogonal to $X_{\underline{t}-\underline{j}}$ via

$$\operatorname{Cov}\Big(X_{\underline{t}} - \sum_{\underline{k} \in \Theta} a_{\underline{k}}\, X_{\underline{t}-\underline{k}}\,,\, X_{\underline{t}-\underline{j}} \Big) = \operatorname{Cov}\Big(\varepsilon_{\underline{t}}\,,\, \varepsilon_{\underline{t}-\underline{j}} + \sum_{\underline{k} \in \Theta} b_{\underline{k}}\, \varepsilon_{\underline{t}-\underline{j}-\underline{k}} \Big) = 0,$$

since $(\varepsilon_{\underline{t}})$ is white noise. It remains to show that the coefficients in (2.5) are uniquely determined. Assume there was a different sequence of coefficients $(a'_{\underline{k}})_{\underline{k} \in \Theta}$ such that

$$X_{\underline{t}} = \sum_{\underline{k} \in \Theta} a'_{\underline{k}}\, X_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}}.$$

Then we get from (2.5) $\sum_{\underline{k}\in\Theta}(a_{\underline{k}} - a'_{\underline{k}})\, X_{\underline{t}-\underline{k}} = 0$. If there exists $\underline{s} \in \Theta$ such that $a_{\underline{s}} \neq a'_{\underline{s}}$ it follows

$$X_{\underline{t}-\underline{s}} = - \sum_{\underline{k}\in\Theta,\, \underline{k}\neq\underline{s}} \frac{a_{\underline{k}} - a'_{\underline{k}}}{a_{\underline{s}} - a'_{\underline{s}}}\, X_{\underline{t}-\underline{k}} \in \overline{sp}\{X_{\underline{j}},\, \underline{j} \neq \underline{t} - \underline{s}\},$$

which contradicts the basic process condition from Assumption 1. Therefore, the coefficients $(a_{\underline{k}})$ are unique. Analogously, assume the MA representation in (2.5) was also fulfilled with coefficients $(b'_{\underline{k}})$, $b_{\underline{s}} \neq b'_{\underline{s}}$. Then we get

$$\varepsilon_{\underline{t}-\underline{s}} = - \sum_{\underline{k}\in\Theta,\, \underline{k}\neq\underline{s}} \frac{b_{\underline{k}} - b'_{\underline{k}}}{b_{\underline{s}} - b'_{\underline{s}}}\, \varepsilon_{\underline{t}-\underline{k}}.$$

Since $f_\varepsilon(\underline{\lambda}) > 0$ implies $\mathrm{Var}(\varepsilon_{\underline{t}}) > 0$ this yields a contradiction via

$$0 < \mathrm{Cov}(\varepsilon_{\underline{t}-\underline{s}}, \varepsilon_{\underline{t}-\underline{s}}) = \mathrm{Cov}\Big(\varepsilon_{\underline{t}-\underline{s}}\,,\, - \sum_{\underline{k}\in\Theta,\, \underline{k}\neq\underline{s}} \frac{b_{\underline{k}} - b'_{\underline{k}}}{b_{\underline{s}} - b'_{\underline{s}}}\, \varepsilon_{\underline{t}-\underline{k}}\Big) = 0,$$

as $(\varepsilon_{\underline{t}})$ is white noise. Hence, the coefficients $(b_{\underline{k}})$ are uniquely determined. $\square$

Proof of Lemma 2.3:

For any $r \geq 0$ and arbitrary functions $g, h \in C_r$ the Fourier series of gh is given by $\sum_{\underline{k}\in\mathbb{Z}^2}(\sum_{\underline{j}\in\mathbb{Z}^2} \widetilde{g}_{\underline{j}}\widetilde{h}_{\underline{k}-\underline{j}})\, e^{i\langle\underline{k},\underline{\lambda}\rangle}$, and from

$$\begin{aligned}(1 + |\underline{k}|_\infty)^r &\leq (1 + |\underline{j}|_\infty + |\underline{k} - \underline{j}|_\infty + |\underline{j}|_\infty \cdot |\underline{k} - \underline{j}|_\infty)^r \\ &= (1 + |\underline{j}|_\infty)^r \cdot (1 + |\underline{k} - \underline{j}|_\infty)^r\end{aligned}$$

one can easily see that $\|\cdot\|_r$ is submultiplicative, i.e.

$$\|gh\|_r \leq \|g\|_r \cdot \|h\|_r. \tag{2.62}$$

Since $f \in C_r$ for $r \geq 2$, its formal Fourier series $\sum_{\underline{k}\in\mathbb{Z}^2} \widetilde{f}_{\underline{k}}\, e^{i\langle\underline{k},\underline{\lambda}\rangle}$ converges absolutely and is therefore equal to $f(\underline{\lambda})$ everywhere on $(-\pi, \pi]^2$. Also, f is twice continuously differentiable and the derivatives are equal to their respective Fourier series

$$\frac{\partial f}{\partial \lambda_j}(\underline{\lambda}) = \sum_{\underline{k}\in\mathbb{Z}^2} ik_j\, \widetilde{f}_{\underline{k}}\, e^{i\langle\underline{k},\underline{\lambda}\rangle}, \qquad \frac{\partial^2 f}{\partial \lambda_1 \partial \lambda_2}(\underline{\lambda}) = \sum_{\underline{k}\in\mathbb{Z}^2} (-k_1 k_2)\, \widetilde{f}_{\underline{k}}\, e^{i\langle\underline{k},\underline{\lambda}\rangle}$$

as well, since these series are obviously absolutely convergent. To be more precise, one has from $|k_1| \cdot |k_2| \leq |\underline{k}|_\infty^2$ that

$$\left\|\frac{\partial^2 f}{\partial \lambda_1 \partial \lambda_2}\right\|_{r-2} = \sum_{\underline{k}\in\mathbb{Z}^2} (1 + |\underline{k}|_\infty)^{r-2}\, |k_1|\, |k_2|\, |\widetilde{f}_{\underline{k}}| \leq \sum_{\underline{k}\in\mathbb{Z}^2} (1 + |\underline{k}|_\infty)^r\, |\widetilde{f}_{\underline{k}}| < \infty,$$

i.e. the second order derivative is in C_{r-2}. The same holds true for the first order derivatives, and Theorem 6.2 in Gröchenig (2007) (the weight function $(1+|\underline{k}|_\infty)^r$ obviously fulfils the required GRS-condition) implies $\|1/f\|_r < \infty$ and in particular $(1/f) \in C_{r-2}$. Since $f(\underline{\lambda}) \geq c > 0$, $\log f$ is also twice continuously differentiable and it holds

$$\frac{\partial^2 \log f}{\partial\lambda_1\partial\lambda_2}(\underline{\lambda}) = \frac{1}{f^2(\underline{\lambda})} \cdot \Big(\frac{\partial^2 f}{\partial\lambda_1\partial\lambda_2}(\underline{\lambda}) \cdot f(\underline{\lambda}) - \frac{\partial f}{\partial\lambda_1}(\underline{\lambda}) \cdot \frac{\partial f}{\partial\lambda_2}(\underline{\lambda})\Big).$$

With this representation, (2.62) and the results established so far we can now infer

$$\Big\|\frac{\partial^2 \log f}{\partial\lambda_1\partial\lambda_2}\Big\|_{r-2} \leq \Big\|\frac{\partial^2 f}{\partial\lambda_1\partial\lambda_2}\Big\|_{r-2} \cdot \Big\|\frac{1}{f}\Big\|_{r-2} + \Big\|\frac{\partial f}{\partial\lambda_1}\Big\|_{r-2} \cdot \Big\|\frac{\partial f}{\partial\lambda_2}\Big\|_{r-2} \cdot \Big\|\frac{1}{f}\Big\|_{r-2}^2 < \infty.$$

The same holds true for the first order derivatives $\partial \log f/\partial\lambda_j$. Now let $(d_{\underline{k}})$ be the Fourier coefficients of $\log f$. As seen above, $\partial \log f/\partial\lambda_j$ and $\partial^2 \log f/\partial\lambda_1\partial\lambda_2$ have Fourier coefficients $ik_j\, d_{\underline{k}}$ and $-k_1k_2\, d_{\underline{k}}$, respectively. Hence, it holds

$$\sum_{\underline{k}\in\mathbb{Z}^2} (1+|\underline{k}|_\infty)^{r-2}\, |k_1|\, |k_2|\, |d_{\underline{k}}| < \infty, \;\sum_{\underline{k}\in\mathbb{Z}^2} (1+|\underline{k}|_\infty)^{r-2}\, |k_j|\, |d_{\underline{k}}| < \infty, \; j=1,2. \tag{2.63}$$

For $k_1, k_2 \in \mathbb{Z}\setminus\{0\}$ it holds $1+|\underline{k}|_\infty \leq 2\,|\underline{k}|_\infty \leq 2\,|k_1|\,|k_2|$. Analogously, $1+|\underline{k}|_\infty$ can be bounded from above by $2\,|k_1|$ if $k_1 \neq 0$, $k_2 = 0$ and by $2\,|k_2|$ if $k_2 \neq 0$, $k_1 = 0$. Therefore, we obtain

$$\begin{aligned} \|\log f\|_{r-1} \leq |d_{\underline{0}}| \;&+\; \sum_{k_1\neq 0} (1+|(k_1,0)'|_\infty)^{r-2}\, 2\,|k_1|\,|d_{\underline{k}}| \\ &+\; \sum_{k_2\neq 0} (1+|(0,k_2)'|_\infty)^{r-2}\, 2\,|k_2|\,|d_{\underline{k}}| \\ &+\; \sum_{k_1,k_2\neq 0} (1+|\underline{k}|_\infty)^{r-2}\, 2\,|k_1|\,|k_2|\,|d_{\underline{k}}| \end{aligned}$$

which is finite due to (2.63). This completes the proof of assertion (i). Assertion (ii) can be proven with analogous arguments for all $r_1, r_2 \geq 1$. □

Proof of Lemma 2.6:

As a preliminary consideration we recall for the vectors $\underline{k}_1, \ldots, \underline{k}_{\bar{p}}$ from (2.9) and arbitrary $r, s \in \{1, \ldots, \bar{p}\}$

$$\int_{(-\pi,\pi]^2} \exp(i\langle \underline{k}_r - \underline{k}_s, \underline{\lambda}\rangle)\, d\underline{\lambda} = \begin{cases} 4\pi^2 & , \; r = s \\ 0 & , \; r \neq s \end{cases}, \tag{2.64}$$

because $\underline{k}_r = \underline{k}_s$ if and only if $r = s$. Let $\underline{d} \in \mathbb{R}^{\bar{p}}$ be arbitrary with $\underline{d} \neq \underline{0}$ and denote by $\underline{w}(\underline{\lambda}) := (\exp(i\langle \underline{k}_1, \underline{\lambda}\rangle), \ldots, \exp(i\langle \underline{k}_{\bar{p}}, \underline{\lambda}\rangle))'$. Observe that $|\underline{d}'\underline{w}(\underline{\lambda})|^2 = \sum_{r,s=1}^{\bar{p}} d_r\, d_s \exp(i\langle \underline{k}_r - \underline{k}_s, \underline{\lambda}\rangle)$. Using (2.64) as well as $\gamma(\underline{h}) = \int_{(-\pi,\pi]^2} f(\underline{\lambda})\, e^{i\langle \underline{h}, \underline{\lambda}\rangle}\, d\underline{\lambda}$ and $f(\underline{\lambda}) \geq c > 0$, cf. Assumption 1, we can derive

$$\begin{aligned}
\underline{d}'\,\Gamma(p)\,\underline{d} &= \int_{(-\pi,\pi]^2} f(\lambda)\, |\underline{d}'\underline{w}(\underline{\lambda})|^2\, d\underline{\lambda} \\
&\geq c \cdot \int_{(-\pi,\pi]^2} |\underline{d}'\underline{w}(\underline{\lambda})|^2\, d\underline{\lambda} \\
&= c \cdot \sum_{r,s=1}^{\bar{p}} d_r\, d_s \int_{(-\pi,\pi]^2} \exp(i\langle \underline{k}_r - \underline{k}_s, \underline{\lambda}\rangle)\, d\underline{\lambda} \\
&= 4\pi^2 c \cdot \underline{d}'\, \underline{d}.
\end{aligned}$$

On the one hand this shows that $\Gamma(p)$ is positive definite and therefore invertible for each $p \in \mathbb{N}$. On the other hand it follows

$$\frac{\underline{d}'\,\Gamma(p)\,\underline{d}}{\underline{d}'\,\underline{d}} \geq 4\pi^2 c,$$

which implies for the smallest eigenvalue $\sigma_{\min}(\Gamma(p)) \geq 4\pi^2 c$, cf. Lütkepohl (1996), 5.2.2 (2). This yields for the largest eigenvalue of the inverse matrix $\sigma_{\max}(\Gamma(p)^{-1}) \leq (4\pi^2 c)^{-1}$ for all $p \in \mathbb{N}$. The spectral norm of the symmetric matrix $\Gamma(p)^{-1}$ is given by its largest eigenvalue, i.e. $\|\Gamma(p)^{-1}\|_{\text{spec}} \leq (4\pi^2 c)^{-1}$ for all $p \in \mathbb{N}$, which yields the desired assertion. $\square$

Proof of Lemma 2.8:

Let $p \in \mathbb{N}$ be arbitrary. For any $\underline{z} = (z_1, z_2) \in S_p$ we define $\underline{\tilde{z}} = (\tilde{z}_1, \tilde{z}_2)$ as the unique vector in S that minimizes the distance to $\underline{z}$ componentwise. To be more precise, let

$$\tilde{z}_1 := \underset{|u_1|=1}{\arg\min}\, |u_1 - z_1|, \quad \tilde{z}_2 := \underset{|u_2|\leq 1}{\arg\min}\, |u_2 - z_2|.$$

In the first step we derive an expression $D(p)$ such that

$$\sup_{\underline{z} \in S_p} \left| A_p(\underline{z}) - A(\underline{\tilde{z}}) \right| \leq D(p). \tag{2.65}$$

For any $\underline{z} \in S_p$ we have

$$\begin{aligned}
\left| A_p(\underline{z}) - A(\underline{\tilde{z}}) \right| &\leq \sum_{\underline{k} \in \Theta(p)} \left| a_{\underline{k}}(p)\, z_1^{k_1}\, z_2^{k_2} - a_{\underline{k}}\, \tilde{z}_1^{k_1}\, \tilde{z}_2^{k_2} \right| + \sum_{\underline{k} \in \Theta \setminus \Theta(p)} |a_{\underline{k}}|\, |\tilde{z}_1|^{k_1}\, |\tilde{z}_2|^{k_2} \\
&\leq \sum_{\underline{k} \in \Theta(p)} |a_{\underline{k}}(p)| \left| z_1^{k_1}\, z_2^{k_2} - \tilde{z}_1^{k_1}\, \tilde{z}_2^{k_2} \right| + \sum_{\underline{k} \in \Theta(p)} |a_{\underline{k}}(p) - a_{\underline{k}}|\, |\tilde{z}_1|^{k_1}\, |\tilde{z}_2|^{k_2}
\end{aligned}$$

$$+\sum_{\underline{k}\in\Theta\setminus\Theta(p)} |a_{\underline{k}}|\,|\widetilde{z}_1|^{k_1}\,|\widetilde{z}_2|^{k_2},$$

which implies

$$\begin{aligned}\sup_{\underline{z}\in S_p}\Big|A_p(\underline{z})-A(\widetilde{\underline{z}})\Big| \;&\le\; \sum_{\underline{k}\in\Theta(p)}|a_{\underline{k}}(p)|\,\sup_{\underline{z}\in S_p}\Big|z_1^{k_1}\,z_2^{k_2}-\widetilde{z}_1^{k_1}\,\widetilde{z}_2^{k_2}\Big| \\ &\quad+\sum_{\underline{k}\in\Theta(p)}|a_{\underline{k}}(p)-a_{\underline{k}}|+\sum_{\underline{k}\in\Theta\setminus\Theta(p)}|a_{\underline{k}}|, \end{aligned}\tag{2.66}$$

since $|\widetilde{z}_1|^{k_1}=1$ and $|\widetilde{z}_2|^{k_2}\le 1$ for any $\underline{k}\in\Theta(p)$. In order to get a bound for the remaining supremum on the right-hand side, consider the following: For arbitrary $\underline{z}\in S_p$, if $|z_2|\le 1$, it follows per definition $\widetilde{z}_2=z_2$, and thus $|z_2^{k_2}-\widetilde{z}_2^{k_2}|=0$. However, if $|z_2|>1$, we can write $z_2=r\,e^{i\varphi}$ for some $-\pi<\varphi\le\pi$ and some $1<r\le(p+1)/p$. For this z_2, it holds $\widetilde{z}_2=e^{i\varphi}$ and therefore

$$\Big|z_2^{k_2}-\widetilde{z}_2^{k_2}\Big|=\Big|r^{k_2}\,e^{ik_2\varphi}-e^{ik_2\varphi}\Big|=r^{k_2}-1\le\Big(\frac{p+1}{p}\Big)^{k_2}-1.$$

Similarly, one can show that

$$\Big|z_1^{k_1}-\widetilde{z}_1^{k_1}\Big|\le\Big(\frac{p+1}{p}\Big)^{|k_1|}-1,$$

for any $\underline{z}\in S_p$, which yields

$$\begin{aligned}\sup_{\underline{z}\in S_p}\Big|z_1^{k_1}\,z_2^{k_2}-\widetilde{z}_1^{k_1}\,\widetilde{z}_2^{k_2}\Big| \;&\le\; \sup_{\underline{z}\in S_p}\Big(\Big|z_1^{k_1}\Big|\Big|z_2^{k_2}-\widetilde{z}_2^{k_2}\Big|+\Big|\widetilde{z}_2^{k_2}\Big|\Big|z_1^{k_1}-\widetilde{z}_1^{k_1}\Big|\Big)\\ &\le\; \Big(\frac{p+1}{p}\Big)^{|k_1|}\cdot\bigg(\Big(\frac{p+1}{p}\Big)^{k_2}-1+\Big(\frac{p+1}{p}\Big)^{|k_1|}-1\bigg).\end{aligned}$$

Inserting this inequality into (2.66) yields (2.65) with

$$\begin{aligned}D(p)\;&:=\;\sum_{\underline{k}\in\Theta(p)}|a_{\underline{k}}(p)|\Big(\frac{p+1}{p}\Big)^{|k_1|}\cdot\bigg(\Big(\frac{p+1}{p}\Big)^{k_2}+\Big(\frac{p+1}{p}\Big)^{|k_1|}-2\bigg)\\ &\quad+\sum_{\underline{k}\in\Theta(p)}|a_{\underline{k}}(p)-a_{\underline{k}}|+\sum_{\underline{k}\in\Theta\setminus\Theta(p)}|a_{\underline{k}}|.\end{aligned}$$

In the next step we show $D(p)\to 0$, as $p\to\infty$. In order to handle the first summand in the definition of $D(p)$, consider that it holds for any $p\in\mathbb{N}$ and any $\underline{k}\in\Theta(p)$

$$\bigg|\Big(\frac{p+1}{p}\Big)^{|k_1|}\cdot\bigg(\Big(\frac{p+1}{p}\Big)^{k_2}+\Big(\frac{p+1}{p}\Big)^{|k_1|}-2\bigg)\bigg|\le 4\Big(\frac{p+1}{p}\Big)^{2p}\le 4\,e^2.\tag{2.67}$$

This, together with $|a_{\underline{k}}(p)| \leq |a_{\underline{k}}| + |a_{\underline{k}}(p) - a_{\underline{k}}|$, yields

$$\begin{aligned} D(p) \leq & \sum_{\underline{k} \in \Theta(p)} |a_{\underline{k}}| \left(\frac{p+1}{p}\right)^{|k_1|} \cdot \left(\left(\frac{p+1}{p}\right)^{k_2} + \left(\frac{p+1}{p}\right)^{|k_1|} - 2 \right) \\ & + (4\,e^2 + 1) \sum_{\underline{k} \in \Theta(p)} |a_{\underline{k}}(p) - a_{\underline{k}}| + \sum_{\underline{k} \in \Theta \setminus \Theta(p)} |a_{\underline{k}}|. \end{aligned} \tag{2.68}$$

For the latter two summands on the right-hand side we immediately get from Theorem 2.7 for some constant $C < \infty$

$$(4\,e^2 + 1) \sum_{\underline{k} \in \Theta(p)} |a_{\underline{k}}(p) - a_{\underline{k}}| + \sum_{\underline{k} \in \Theta \setminus \Theta(p)} |a_{\underline{k}}| \leq \left(C(4\,e^2 + 1) + 1 \right) \sum_{\underline{k} \in \Theta \setminus \Theta(p)} (1 + |\underline{k}|_\infty)|a_{\underline{k}}|,$$

which converges to zero, as $p \to \infty$, due to summability condition (2.6) and since $\Theta(p) \to \Theta$. For the first summand on the right-hand side of inequality (2.68) we can apply Lebesgue's dominated convergence theorem because (2.67) provides a dominating and summable sequence via

$$\begin{aligned} & \sum_{\underline{k} \in \Theta} \mathbb{1}_{\{\underline{k} \in \Theta(p)\}} |a_{\underline{k}}| \left| \left(\frac{p+1}{p}\right)^{|k_1|} \cdot \left(\left(\frac{p+1}{p}\right)^{k_2} + \left(\frac{p+1}{p}\right)^{|k_1|} - 2 \right) \right| \\ \leq \; & 4\,e^2 \sum_{\underline{k} \in \Theta} |a_{\underline{k}}| < \infty. \end{aligned}$$

Hence,

$$\begin{aligned} & \lim_{p \to \infty} \sum_{\underline{k} \in \Theta} \mathbb{1}_{\{\underline{k} \in \Theta(p)\}} |a_{\underline{k}}| \left(\frac{p+1}{p}\right)^{|k_1|} \left(\left(\frac{p+1}{p}\right)^{k_2} + \left(\frac{p+1}{p}\right)^{|k_1|} - 2 \right) \\ = \; & \sum_{\underline{k} \in \Theta} |a_{\underline{k}}| \lim_{p \to \infty} \left(\frac{p+1}{p}\right)^{|k_1|} \left(\left(\frac{p+1}{p}\right)^{k_2} + \left(\frac{p+1}{p}\right)^{|k_1|} - 2 \right) \\ = \; & 0. \end{aligned}$$

Therefore, we have $D(p) \to 0$, as $p \to \infty$. From the representation of $A(\underline{\tilde{z}})$ as an exponential of a bounded function, cf. (2.61) and (2.60), we have $|A(\underline{\tilde{z}})| \geq 2\delta$ uniformly for all $\underline{\tilde{z}} \in S$ and for some $\delta > 0$. Then, choosing p large enough such that $D(p) \leq \delta$, (2.65) implies

$$|A_p(\underline{z})| \geq \delta \quad \forall \underline{z} \in S_p, \tag{2.69}$$

which is the first assertion of Lemma 2.8.

Now let p be large enough such that (2.69) holds, but fixed. We will derive the series representation of $B_p(\underline{z}) = 1/A_p(\underline{z})$ as in Lemma 2.8. Using the equivalent notations $a_{\underline{k}}(p)$ and $a_{(k_1,k_2)}(p)$ for the coefficients, we can write

$$A_p(\underline{z}) = 1 + \sum_{\underline{k}\in\Theta(p)} a_{\underline{k}}(p)\, z_1^{k_1} z_2^{k_2} = \sum_{k_2=0}^{p} \alpha(z_1,k_2)\, z_2^{k_2},$$

where

$$\alpha(z_1,k_2) := \sum_{k_1=-p}^{p} a_{(k_1,k_2)}(p)\, z_1^{k_1},$$

with $a_{(0,0)}(p) := 1$ and $a_{(k_1,0)}(p) := 0$ for all $k_1 < 0$. Let z_1 be fixed with $p/(p+1) \le |z_1| \le (p+1)/p$. Then, since $A_p(z_1,z_2)$ is a polynomial in z_2 (for each fixed z_1) with a finite number of complex roots, which is bounded away from zero on $|z_2| \le (p+1)/p$, it actually has no complex roots on a slightly larger open disk $|z_2| < (p+1)/p + \varepsilon$, and $B_p(z_1,z_2) = 1/A_p(z_1,z_2)$ is an analytic function (in z_2) on this open disk $|z_2| < (p+1)/p + \varepsilon$. Thus, $B_p(z_1,z_2)$ can be represented as a power series (in z_2)

$$B_p(z_1,z_2) = \Big(\sum_{k_2=0}^{p} \alpha(z_1,k_2)\, z_2^{k_2}\Big)^{-1} = \sum_{k_2=0}^{\infty} \beta(z_1,k_2)\, z_2^{k_2}, \tag{2.70}$$

which converges absolutely on $|z_2| \le (p+1)/p$. The coefficients $\beta(z_1,k_2)$ can be determined recursively from

$$1 = \sum_{k_2=0}^{p} \alpha(z_1,k_2)\, z_2^{k_2} \cdot \sum_{k_2=0}^{\infty} \beta(z_1,k_2)\, z_2^{k_2} = \sum_{k_2=0}^{\infty} \Big(\sum_{l_2=0}^{p\wedge k_2} \alpha(z_1,l_2)\, \beta(z_1,k_2-l_2)\Big) z_2^{k_2}.$$

For example, if $p \ge 2$, one can derive

$$\beta(z_1,0) = \frac{1}{\alpha(z_1,0)},\quad \beta(z_1,1) = \frac{-\alpha(z_1,1)}{\alpha(z_1,0)^2},\quad \beta(z_1,2) = \frac{\alpha(z_1,1)^2 - \alpha(z_1,0)\alpha(z_1,2)}{\alpha(z_1,0)^3},$$

and so on. In general, one can obtain that

$$\beta(z_1,k_2) = \frac{\eta(z_1,k_2)}{\alpha(z_1,0)^{k_2+1}}, \tag{2.71}$$

where $\eta(z_1,k_2)$ is some finite linear combination of certain k_2-fold products of the coefficients $\alpha(z_1,0),\dots,\alpha(z_1,p)$. Hence, it is easy to see that each $\eta(z_1,k_2)$ can be expressed as

$$\eta(z_1,k_2) = \sum_{k_1=-pk_2}^{pk_2} c_{(k_1,k_2)}(p)\, z_1^{k_1}, \tag{2.72}$$

defined on $p/(p+1) \leq |z_1| \leq (p+1)/p$, for suitable coefficients $c_{(k_1,k_2)}(p)$.
We will now develop Laurent series expansions (in z_1) for each $\beta(z_1, k_2)$. At first, observe that in (2.10) the z-transform $A(z_1, z_2)$ was defined on the domain S, i.e. on $|z_1| = 1$, $|z_2| \leq 1$, because the series converges absolutely on S due to

$$|A(z_1, z_2)| \leq 1 + \sum_{\underline{k} \in \Theta} |a_{(k_1,k_2)}| \left|z_1^{k_1}\right| \left|z_2^{k_2}\right| \leq 1 + \sum_{\underline{k} \in \Theta} |a_{(k_1,k_2)}| < \infty.$$

Note that in the series expansion in (2.10), only exponents $k_2 \geq 0$ but both positive and negative exponents k_1 show up. However, for $z_2 = 0$ fixed, the series reduces to

$$A(z_1, 0) = 1 + \sum_{k_1=1}^{\infty} a_{(k_1,0)}\, z_1^{k_1}$$

with only positive exponents k_1. Therefore the series expansion $A(z_1, 0)$ actually converges absolutely not only on the unit circle $|z_1| = 1$, but on the entire disk $|z_1| \leq 1$. Analogously, $A_p(z_1, 0)$ reduces to a polynomial

$$A_p(z_1, 0) = 1 + \sum_{k_1=1}^{p} a_{(k_1,0)}(p)\, z_1^{k_1},$$

which is defined not only on the ring $p/(p+1) \leq |z_1| \leq (p+1)/p$ but on the closed disk $|z_1| \leq (p+1)/p$. Again, from the representation of $A(z_1, 0)$ as an exponential of a bounded function, cf. (2.61) and (2.60), we get that $|A(z_1, 0)|$ is uniformly bounded away from zero on $|z_1| \leq 1$. Also, with the very same technique as for showing (2.65), we can derive

$$\sup_{|z_1| \leq 1} \left|A_p(z_1, 0) - A(z_1, 0)\right| \leq D(p),$$

and therefore, for the fixed p large enough chosen above, we have

$$|A_p(z_1, 0)| \geq \delta \quad \forall\, |z_1| \leq (p+1)/p.$$

Hence $1/A_p(z_1, 0)$ can be expanded as an absolutely convergent power series on $|z_1| \leq (p+1)/p$. Since we also have per definition $\alpha(z_1, 0) = A_p(z_1, 0)$ and $\beta(z_1, 0) = 1/\alpha(z_1, 0)$, it holds

$$\beta(z_1, 0) = \frac{1}{A_p(z_1, 0)} = 1 + \sum_{k_1=1}^{\infty} b_{(k_1,0)}(p)\, z_1^{k_1},$$

for suitable coefficients $b_{(k_1,0)}(p)$. It follows immediately that for each $k_2 \geq 1$

$$\frac{1}{\alpha(z_1, 0)^{k_2+1}} = \beta(z_1, 0)^{k_2+1} = 1 + \sum_{k_1=1}^{\infty} \widetilde{b}_{(k_1,k_2)}(p)\, z_1^{k_1},$$

for suitable coefficients $\widetilde{b}_{(k_1,k_2)}(p)$, the series absolutely convergent on $|z_1| \le (p+1)/p$. This expansion, together with (2.72) and (2.71), shows that for all $k_2 \ge 1$

$$\beta(z_1, k_2) = \sum_{k_1=-pk_2}^{\infty} b_{(k_1,k_2)}(p)\, z_1^{k_1},$$

absolutely convergent on $p/(p+1) \le |z_1| \le (p+1)/p$, for suitable coefficients $b_{(k_1,k_2)}(p)$. Inserting this into (2.70), and setting $b_{(k_1,k_2)}(p) := 0$ for all $k_1 < -k_2 p$, yields

$$\begin{aligned} B_p(z_1, z_2) &= \sum_{k_2=0}^{\infty} \beta(z_1, k_2)\, z_2^{k_2} \\ &= 1 + \sum_{k_1=1}^{\infty} b_{(k_1,0)}(p)\, z_1^{k_1} + \sum_{k_2=1}^{\infty} \sum_{k_1=-pk_2}^{\infty} b_{(k_1,k_2)}(p)\, z_1^{k_1}\, z_2^{k_2} \\ &= 1 + \sum_{\underline{k}\in\Theta} b_{\underline{k}}(p)\, z_1^{k_1}\, z_2^{k_2}, \end{aligned}$$

which completes the proof. □

Proof of Lemma 2.9:

We will use the space C_r of functions on $(-\pi, \pi]^2$ with finite norm $\|\cdot\|_r$ as defined in Lemma 2.3. In the proof of Lemma 2.1 we introduced the functions $A', B' \in C_{r-1}$ which possess the representations

$$A'(\underline{\lambda}) = \widetilde{a}_{\underline{0}}\Big(1 - \sum_{\underline{k}\in\Theta} a_{\underline{k}}\, e^{-i\langle \underline{k},\underline{\lambda}\rangle}\Big), \quad B'(\underline{\lambda}) = \widetilde{b}_{\underline{0}}\Big(1 + \sum_{\underline{k}\in\Theta} b_{\underline{k}}\, e^{-i\langle \underline{k},\underline{\lambda}\rangle}\Big),$$

respectively, as can be seen from (2.61). Hence, the Fourier coefficients of A', B' are given by the autoregressive and moving average parameters $a_{\underline{k}}$ and $b_{\underline{k}}$, up to the constant non-zero factors $\widetilde{a}_{\underline{0}}, \widetilde{b}_{\underline{0}}$. In order to simplify the notation in the remainder of this proof we define for all $p \ge p_0$ the functions $A'_p, B'_p : (-\pi, \pi]^2 \to \mathbb{R}$ via $A'_p(\underline{\lambda}) := \widetilde{a}_{\underline{0}}\, A_p(e^{-i\lambda_1}, e^{-i\lambda_2})$, $B'_p(\underline{\lambda}) := \widetilde{b}_{\underline{0}}\, B_p(e^{-i\lambda_1}, e^{-i\lambda_2})$ and obtain

$$A'_p(\underline{\lambda}) = \widetilde{a}_{\underline{0}}\Big(1 - \sum_{\underline{k}\in\Theta(p)} a_{\underline{k}}(p)\, e^{-i\langle \underline{k},\underline{\lambda}\rangle}\Big), \quad B'_p(\underline{\lambda}) = \widetilde{b}_{\underline{0}}\Big(1 + \sum_{\underline{k}\in\Theta} b_{\underline{k}}(p)\, e^{-i\langle \underline{k},\underline{\lambda}\rangle}\Big),$$

cf. (2.11) and (2.12) for the definitions of $A_p(\underline{z})$ and $B_p(\underline{z})$. Since $\widetilde{a}_{\underline{0}} = 1/\widetilde{b}_{\underline{0}}$, we can conclude from (2.60) and (2.12) that

$$A'_p(\underline{\lambda}) = 1/B'_p(\underline{\lambda}), \quad A'(\underline{\lambda}) = 1/B'(\underline{\lambda}) \quad \forall\, \underline{\lambda} \in (-\pi, \pi]^2. \tag{2.73}$$

We now have established the necessary notation to prove the assertion of Lemma 2.9. For all $s \in \mathbb{N}_0$ with $s+1 < r$ we derive, using (2.73) and the submultiplicativity of $\|\cdot\|_s$ established in (2.62),

$$\begin{aligned}
&\sum_{k\in\Theta}(1+|\underline{k}|_\infty)^s\,|b_{\underline{k}}(p)-b_{\underline{k}}| \\
=\;& \Big\|\,(1/\widetilde{b}_{\underline{0}})\cdot(B_p'-B')\,\Big\|_s \qquad (2.74)\\
=\;& (1/\widetilde{b}_{\underline{0}})\cdot\Big\|B_p'\cdot\Big[A'-A_p'\Big]\cdot B'\Big\|_s \\
\le\;& (1/\widetilde{b}_{\underline{0}})\cdot\Big(\|B_p'-B'\|_s+\|B'\|_s\Big)\cdot\|A'-A_p'\|_s\cdot\|B'\|_s. \qquad (2.75)
\end{aligned}$$

From Baxter's inequality, cf. Theorem 2.7, we can infer

$$\begin{aligned}
\|A'-A_p'\|_s \;&=\; \widetilde{a}_{\underline{0}}\cdot\Bigg(\sum_{k\in\Theta(p)}(1+|\underline{k}|_\infty)^s\,|a_{\underline{k}}(p)-a_{\underline{k}}|+\sum_{k\in\Theta\setminus\Theta(p)}(1+|\underline{k}|_\infty)^s\,|a_{\underline{k}}|\Bigg)\\
&\le\; \widetilde{a}_{\underline{0}}\Big(\frac{M}{2\sqrt{2}\,\pi^2 c}+1\Big)\cdot\sum_{k\in\Theta\setminus\Theta(p)}(1+|\underline{k}|_\infty)^{s+1}\,|a_{\underline{k}}| \qquad \forall\, p\ge p_0. \qquad (2.76)
\end{aligned}$$

Because the right-hand side converges to zero as $p\to\infty$, one can always find $p\in\mathbb{N}$ such that $\|A'-A_p'\|_s$ becomes arbitrarily small. In particular, for some arbitrary $\delta\in(0,1)$, choose $p_1\ge p_0$ such that

$$\|A'-A_p'\|_s\cdot\|B'\|_s\le\delta$$

for all $p\ge p_1$. Taking the difference of (2.75) and (2.74) we get

$$\begin{aligned}
\|B_p'-B'\|_s \;&\le\; \frac{\|B'\|_s^2\cdot\|A'-A_p'\|_s}{1-\|A'-A_p'\|_s\cdot\|B'\|_s}\\
&\le\; \frac{\|B'\|_s^2}{1-\delta}\cdot\|A'-A_p'\|_s \qquad (2.77)
\end{aligned}$$

for all $p\ge p_2$. Since the first factor on the right-hand side of (2.77) does not depend on p and is finite, applying (2.76) to the second factor yields that there exists $C<\infty$ such that

$$\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^s\,|b_{\underline{k}}(p)-b_{\underline{k}}|\le C\cdot\sum_{\underline{k}\in\Theta\setminus\Theta(p)}(1+|\underline{k}|_\infty)^{s+1}\,|a_{\underline{k}}| \qquad \forall\, p\ge p_2,$$

which completes the proof. □

Proof of Lemma 2.10:

Due to Lemma 2.8 and Assumption 2, we can choose $\delta>0$ and $n_0\in\mathbb{N}$ large enough such that

$$\Big|A_p(\underline{z})\Big|\ge\delta,\quad \Big|\widehat{A}_p(\underline{z})\Big|\ge\delta \;\text{ in prob.} \qquad (2.78)$$

for all $n \geq n_0$ and for all $\underline{z} \in S_p$. For those n, $B_p(\underline{z}) = 1 + \sum_{\underline{k}\in\Theta} b_{\underline{k}}(p)\, z_1^{k_1} z_2^{k_2}$ can be expanded as a power series in z_2 with coefficients depending on z_1. These coefficients can then be expanded as Laurent series in z_1, cf. the proof of Lemma 2.8 for a detailed explanation and and introduction of the notation which will also be used in this proof. To be precise, we have

$$B_p(z_1, z_2) = \sum_{k_2=0}^{\infty} \beta(z_1, k_2)\, z_2^{k_2}, \quad \text{where } \beta(z_1, k_2) = \sum_{k_1=-\infty}^{\infty} b_{(k_1,k_2)}(p)\, z_1^{k_1},$$

with $b_{(k_1,0)}(p) = 0$ for $k_1 < 0$ and $b_{(0,0)}(p) = 1$. Following exactly along the lines of the proof of Lemma 2.8, we get an expansion with the very same structure for $\widehat{B}_p(\underline{z}) = 1 + \sum_{\underline{k}\in\Theta} \widehat{b}_{\underline{k}}(p)\, z_1^{k_1} z_2^{k_2}$, in probability, as

$$\widehat{B}_p(z_1, z_2) = \sum_{k_2=0}^{\infty} \widehat{\beta}(z_1, k_2)\, z_2^{k_2}, \quad \text{where } \widehat{\beta}(z_1, k_2) = \sum_{k_1=-\infty}^{\infty} \widehat{b}_{(k_1,k_2)}(p)\, z_1^{k_1},$$

also with $\widehat{b}_{(k_1,0)}(p) = 0$ for $k_1 < 0$ and $\widehat{b}_{(0,0)}(p) = 1$. Then, for any $k_2 \geq 0$, we have the Laurent series expansion

$$\widehat{\beta}(z_1, k_2) - \beta(z_1, k_2) = \sum_{k_1=-\infty}^{\infty} \left(\widehat{b}_{(k_1,k_2)}(p) - b_{(k_1,k_2)}(p)\right) z_1^{k_1}$$

in probability, which converges absolutely on the ring $R_1 := p/(p+1) \leq z_1 \leq (p+1)/p$. Actually, following the same argument as for the function $\beta(z_1, k_2)$ in the proof of Lemma 2.8, the function $\widehat{\beta}(z_1, k_2) - \beta(z_1, k_2)$ is analytic (and the Laurent series expansion thus valid) on a slightly larger open set which contains the closed ring R_1 as a subset. Therefore, Cauchy's inequality for analytic functions yields the following bounds for the coefficients:

$$\begin{aligned}
\left|\widehat{b}_{(k_1,k_2)}(p) - b_{(k_1,k_2)}(p)\right| &\leq \left(\frac{p+1}{p}\right)^{-k_1} \sup_{|z_1|=(p+1)/p} \left|\widehat{\beta}(z_1, k_2) - \beta(z_1, k_2)\right| \quad \forall\, k_1 \geq 0, \\
\left|\widehat{b}_{(k_1,k_2)}(p) - b_{(k_1,k_2)}(p)\right| &\leq \left(\frac{p}{p+1}\right)^{-k_1} \sup_{|z_1|=p/(p+1)} \left|\widehat{\beta}(z_1, k_2) - \beta(z_1, k_2)\right| \quad \forall\, k_1 < 0
\end{aligned}$$

in probability. These two bounds can be combined to obtain

$$\begin{aligned}
&\left|\widehat{b}_{(k_1,k_2)}(p) - b_{(k_1,k_2)}(p)\right| \\
&\leq \left(\frac{p+1}{p}\right)^{-|k_1|} \sup_{p/(p+1)\leq|z_1|\leq(p+1)/p} \left|\widehat{\beta}(z_1, k_2) - \beta(z_1, k_2)\right| \qquad (2.79)
\end{aligned}$$

in probability, for all $k_1 \in \mathbb{Z}$. Then, for any $z_1 \in R_1$, $\widehat{B}_p(z_1, z_2) - B_p(z_1, z_2)$, as a function in z_2, has the power series expansion

$$\widehat{B}_p(z_1, z_2) - B_p(z_1, z_2) = \sum_{k_2=0}^{\infty} \left(\widehat{\beta}(z_1, k_2) - \beta(z_1, k_2)\right) z_2^{k_2},$$

in probability, which converges absolutely on the closed disk $|z_2| \leq (p+1)/p$. Hence, Cauchy's inequality yields the bound

$$\left|\widehat{\beta}(z_1,k_2) - \beta(z_1,k_2)\right| \quad \leq \quad \left(\frac{p+1}{p}\right)^{-k_2} \sup_{|z_2|=(p+1)/p} \left|\widehat{B}_p(z_1,z_2) - B_p(z_1,z_2)\right| \quad \text{in prob.}$$

Inserting this bound into (2.79), and using (2.78), we get

$$\begin{aligned}
&\left|\widehat{b}_{(k_1,k_2)}(p) - b_{(k_1,k_2)}(p)\right| \\
\leq\ & \left(\frac{p+1}{p}\right)^{-|k_1|-k_2} \sup_{p/(p+1)\leq|z_1|\leq(p+1)/p,\ |z_2|=(p+1)/p} \left|\widehat{B}_p(z_1,z_2) - B_p(z_1,z_2)\right| \\
\leq\ & \left(\frac{p+1}{p}\right)^{-|k_1|-k_2} \sup_{\underline{z}\in S_p} \left|\frac{\widehat{A}_p(\underline{z}) - A_p(\underline{z})}{\widehat{A}_p(\underline{z})\, A_p(\underline{z})}\right| \\
\leq\ & \left(\frac{p+1}{p}\right)^{-|k_1|-k_2} \frac{1}{\delta^2} \sup_{\underline{z}\in S_p} \sum_{\underline{j}\in\Theta(p)} \left|\widehat{a}_{\underline{j}}(p) - a_{\underline{j}}(p)\right| |z_1|^{j_1}\, |z_2|^{j_2} \\
\leq\ & \left(\frac{p+1}{p}\right)^{-|k_1|-k_2} \left(\frac{p+1}{p}\right)^{2p} \frac{1}{\delta^2} \sum_{\underline{j}\in\Theta(p)} \left|\widehat{a}_{\underline{j}}(p) - a_{\underline{j}}(p)\right| \\
\leq\ & \left(\frac{p+1}{p}\right)^{-|k_1|-k_2} \frac{1}{p^4} \cdot C
\end{aligned}$$

in probability, for some $C < \infty$, because of Assumption 2 and since $((p+1)/p)^{2p}$ is a sequence bounded by e^2. □

Proof of Lemma 2.13, assertion (2.19):
Throughout this proof, we consider only those n large enough such that $A_p(\underline{z})$ and $\widehat{A}_p(\underline{z})$ are bounded away from zero on S_p, the latter in probability, cf. Lemma 2.10. It holds

$$\begin{aligned}
\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2 \left|\widehat{b}_{\underline{k}}(p)\right| \quad \leq \quad & \sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2\, |b_{\underline{k}}| + \sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2\, |b_{\underline{k}}(p) - b_{\underline{k}}| \\
& + \sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2 \left|\widehat{b}_{\underline{k}}(p) - b_{\underline{k}}(p)\right|.
\end{aligned}$$

The first summand on the right-hand side is finite due to (2.6) while the second summand can be bounded with Lemma 2.9 by

$$\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2\, |b_{\underline{k}}(p) - b_{\underline{k}}| \leq C \cdot \sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^3\, |a_{\underline{k}}|,$$

uniformly for all p (and thus all n). The right-hand side, again, is finite due to (2.6). Hence, the proof can be completed by showing

$$\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^2 \left|\widehat{b}_{\underline{k}}(p) - b_{\underline{k}}(p)\right| = \mathcal{O}_P(1).$$

Due to

$$(1+|\underline{k}|_\infty)^2 \leq 1+3\,|\underline{k}|_\infty^2 \leq 3\,(1+|k_1|^2)+3\,(1+|k_2|^2)$$

it suffices to show

$$\sum_{\underline{k}\in\Theta}(1+|k_j|^2)\left|\widehat{b}_{\underline{k}}(p)-b_{\underline{k}}(p)\right| = \mathcal{O}_P(1), \quad j=1,2. \tag{2.80}$$

Let $j=1$. Lemma 2.10 yields

$$\begin{aligned}
&\sum_{\underline{k}\in\Theta}(1+|k_1|^2)\left|\widehat{b}_{\underline{k}}(p)-b_{\underline{k}}(p)\right| \\
\leq\ & C\cdot\frac{1}{p^4}\sum_{\underline{k}\in\Theta}(1+|k_1|^2)\left(\frac{p+1}{p}\right)^{-|k_1|-k_2} \\
\leq\ & C\cdot\frac{1}{p^4}\sum_{k_1=-\infty}^{\infty}(1+|k_1|^2)\left(\frac{p}{p+1}\right)^{|k_1|}\cdot\sum_{k_2=0}^{\infty}\left(\frac{p}{p+1}\right)^{k_2} \quad \text{in prob.}
\end{aligned} \tag{2.81}$$

For any $|x|<1$, differentiating the geometric series twice yields

$$\sum_{m=0}^{\infty}(m+2)\,(m+1)\,x^m = \frac{2}{(1-x)^3},$$

thus, the right-hand side of (2.81) can be bounded by

$$\begin{aligned}
& C\cdot\frac{1}{p^4}\,2\sum_{k_1=0}^{\infty}(k_1+2)\,(k_1+1)\left(\frac{p}{p+1}\right)^{k_1}\cdot\sum_{k_2=0}^{\infty}\left(\frac{p}{p+1}\right)^{k_2} \\
=\ & C\cdot\frac{4}{p^4}\cdot\left(1-\frac{p}{p+1}\right)^{-3}\cdot\left(1-\frac{p}{p+1}\right)^{-1} \\
=\ & C\cdot\frac{4(p+1)^4}{p^4} = \mathcal{O}(1).
\end{aligned}$$

Since the inequality in (2.81) holds in probability, we have shown (2.80) for $j=1$. The same calculation can be performed for $j=2$, which completes the proof. □

Proof of Lemma 2.13, assertion (2.20):

The random variables $\varepsilon^*_{\underline{t}}$ are, conditionally on the given data, uniformly distributed on $\{\widehat{\varepsilon}_{\underline{s}}(p) : \underline{s}\in\Pi(n,p)\}$. Hence, it holds

$$E^*\Big(|\varepsilon^*_{\underline{t}}|^{2w}\Big) = \frac{1}{|\Pi(n,p)|}\sum_{\underline{s}\in\Pi(n,p)}|\widehat{\varepsilon}_{\underline{s}}(p)|^{2w},$$

and the goal is to show that the right-hand side converges to $E\Big(|\varepsilon_{\underline{t}}|^{2w}\Big) = E\Big(|\varepsilon_{\underline{0}}|^{2w}\Big)$. Because of Assumption 4 it suffices to show

$$\frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \Big(|\widehat{\varepsilon}_{\underline{t}}(p)|^{2w} - |\varepsilon_{\underline{t}}|^{2w}\Big) \xrightarrow{P} 0. \tag{2.82}$$

We have $\widehat{\varepsilon}_{\underline{t}}(p) = \varepsilon'_{\underline{t}}(p) - \overline{\varepsilon}$ with $\varepsilon'_{\underline{t}}(p) = X_{\underline{t}} - \sum_{\underline{k}\in\Theta(p)} \widehat{a}_{\underline{k}}(p)\, X_{\underline{t}-\underline{k}}$ and $\overline{\varepsilon} = (1/|\Pi(n,p)|) \sum_{\underline{t}\in\Pi(n,p)} \varepsilon'_{\underline{t}}(p)$. Thus, with representation (2.5), we have

$$\widehat{\varepsilon}_{\underline{t}}(p) = \varepsilon'_{\underline{t}}(p) - \overline{\varepsilon} = \varepsilon_{\underline{t}} + Q_{\underline{t}} + R_{\underline{t}} - \overline{\varepsilon},$$

where

$$\begin{aligned}
Q_{\underline{t}} &:= \sum_{\underline{k}\in\Theta(p)} \Big(a_{\underline{k}}(p) - \widehat{a}_{\underline{k}}(p)\Big)\, X_{\underline{t}-\underline{k}},\\
R_{\underline{t}} &:= \sum_{\underline{k}\in\Theta(p)} \Big(a_{\underline{k}} - a_{\underline{k}}(p)\Big)\, X_{\underline{t}-\underline{k}} + \sum_{\underline{k}\in\Theta\setminus\Theta(p)} a_{\underline{k}}\, X_{\underline{t}-\underline{k}}.
\end{aligned}$$

Decomposing $|\widehat{\varepsilon}_{\underline{t}}(p)|^{2w} = (\varepsilon_{\underline{t}} + Q_{\underline{t}} + R_{\underline{t}} - \overline{\varepsilon})^{2w}$ with a binomial expansion (with the notation $|\underline{d}| = d_1 + d_2 + d_3 + d_4$ for vectors $\underline{d} \in \mathbb{N}_0^4$), one can easily see that it holds for some $C < \infty$

$$\begin{aligned}
&\left|\frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \Big(|\widehat{\varepsilon}_{\underline{t}}(p)|^{2w} - |\varepsilon_{\underline{t}}|^{2w}\Big)\right|\\
\leq\ & C \cdot \sum_{|\underline{d}|=2w,\ d_1\neq 2w} \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} |\varepsilon_{\underline{t}}|^{d_1}\, |Q_{\underline{t}}|^{d_2}\, |R_{\underline{t}}|^{d_3}\, |\overline{\varepsilon}|^{d_4}\\
\leq\ & C \cdot \sum_{|\underline{d}|=2w,\ d_1\neq 2w} (I)^{d_1/2w}\, (II)^{d_2/2w}\, (III)^{d_3/2w}\, (IV)^{d_4/2w},
\end{aligned}$$

where Hölder's inequality was used in the final step and, moreover,

$$\begin{aligned}
(I) &= \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} |\varepsilon_{\underline{t}}|^{2w},\ (II) = \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} |Q_{\underline{t}}|^{2w},\\
(III) &= \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} |R_{\underline{t}}|^{2w},\ (IV) = \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} |\overline{\varepsilon}|^{2w} = |\overline{\varepsilon}|^{2w}.
\end{aligned}$$

Obviously, (2.82) holds true if we can show

$$(I) = \mathcal{O}_P(1),\ (II) = o_P(1),\ (III) = o_P(1),\ (IV) = o_P(1). \tag{2.83}$$

In the first step we show $(IV_n) = o_P(1)$. From the definition of $\overline{\varepsilon}$ we get

$$|\overline{\varepsilon}| \leq \left|\frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \varepsilon_{\underline{t}}\right| + \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \Big|Q_{\underline{t}}\Big| + \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \Big|R_{\underline{t}}\Big|. \tag{2.84}$$

The first summand on the right-hand side obviously converges to zero in probability because of the WLLN (recall that the random variables $\varepsilon_{\underline{t}}$ are uncorrelated and have mean zero). Considering the second summand, we have from Assumption 1 that $|X_{\underline{t}}| = \mathcal{O}_P(1)$ uniformly for all $\underline{t} \in \mathbb{Z}^2$, and, since $|\Theta(p)| = 2p(p+1)$,

$$\sum_{\underline{k}\in\Theta(p)} \left|X_{\underline{t}-\underline{k}}\right| = 2p(p+1)\,\mathcal{O}_P(1) = \mathcal{O}_P(p^2). \tag{2.85}$$

It follows

$$\left|Q_{\underline{t}}\right| \leq \sum_{\underline{k}\in\Theta(p)} \left|a_{\underline{k}}(p) - \widehat{a}_{\underline{k}}(p)\right| \cdot \sum_{\underline{k}\in\Theta(p)} \left|X_{\underline{t}-\underline{k}}\right| \leq \frac{1}{p^4}\,\mathcal{O}_P(1)\cdot\mathcal{O}_P(p^2) = o_P(1), \tag{2.86}$$

where Assumption 2 was used. Since this bound does not depend on $\underline{t}$, we have

$$\frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \left|Q_{\underline{t}}\right| = o_P(1). \tag{2.87}$$

Now consider the third summand on the right-hand side of (2.84). We will need the following preliminary results: From Theorem 2.7 and summability condition (2.6) (here, we assume $r=4$) we get

$$\begin{aligned} & p^2 \cdot \sum_{\underline{k}\in\Theta(p)} |a_{\underline{k}}(p) - a_{\underline{k}}| \\ \leq\; & C \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} p^2\,(1+|\underline{k}|_\infty)\,|a_{\underline{k}}| \leq C \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1+|\underline{k}|_\infty)^3\,|a_{\underline{k}}| = o(1), \end{aligned}$$

because $p \leq |\underline{k}|_\infty$ for all $\underline{k} \in \Theta\setminus\Theta(p)$, and $\Theta(p)\to\Theta$, as $n\to\infty$. Hence we have $\sum_{\underline{k}\in\Theta(p)} |a_{\underline{k}}(p) - a_{\underline{k}}| = o(p^{-2})$. Moreover, since $|X_{\underline{t}}| = \mathcal{O}_P(1)$ uniformly for all $\underline{t}\in\mathbb{Z}^2$, we have

$$p \sum_{\underline{k}\in\Theta\setminus\Theta(p)} |a_{\underline{k}}|\,|X_{\underline{t}-\underline{k}}| \leq \mathcal{O}_P(1) \sum_{\underline{k}\in\Theta\setminus\Theta(p)} p\,|a_{\underline{k}}| \leq \mathcal{O}_P(1)\cdot\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)\,|a_{\underline{k}}| = \mathcal{O}_P(1),$$

due to (2.6). This implies $\sum_{\underline{k}\in\Theta\setminus\Theta(p)} |a_{\underline{k}}|\,|X_{\underline{t}-\underline{k}}| = \mathcal{O}_P(p^{-1})$. Combining these results and (2.85) we get

$$\begin{aligned} \left|R_{\underline{t}}\right| \;&\leq\; \sum_{\underline{k}\in\Theta(p)} \left|a_{\underline{k}} - a_{\underline{k}}(p)\right| \cdot \sum_{\underline{k}\in\Theta(p)} \left|X_{\underline{t}-\underline{k}}\right| + \sum_{\underline{k}\in\Theta\setminus\Theta(p)} |a_{\underline{k}}|\,|X_{\underline{t}-\underline{k}}| \\ &\leq\; o(p^{-2})\cdot\mathcal{O}_P(p^2) + \mathcal{O}_P(p^{-1}) = o_P(1). \end{aligned} \tag{2.88}$$

Since this bound does not depend on $\underline{t}$, we have

$$\frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \left|R_{\underline{t}}\right| = o_P(1). \tag{2.89}$$

Combining this with (2.84) and (2.85) gives $(IV) = o_P(1)$.
Since the bounds in (2.86) and (2.88) do not depend on $\underline{t}$, it follows immediately

$$\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)} |Q_{\underline{t}}|^{2w} = \mathcal{O}_P(p^{-4w}), \quad \frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)} |R_{\underline{t}}|^{2w} = o_P(1),$$

i.e. $(II) = o_P(1)$ and $(III) = o_P(1)$. Furthermore, Assumption 4 guarantees $(I) = \mathcal{O}_P(1)$, which delivers the final assertion of (2.83) and completes the proof of (2.20).
As a byproduct, we get a result about the empirical means of $(\widehat{\varepsilon}_{\underline{t}} - \varepsilon_{\underline{t}})^2$, which will be needed later on. Using the fact that $|\varepsilon_{\underline{t}}| = \mathcal{O}_P(1)$ uniformly for all $\underline{t} \in \mathbb{Z}^2$, we can derive

$$\begin{aligned}
&\left|\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)} \left(\widehat{\varepsilon}_{\underline{t}} - \varepsilon_{\underline{t}}\right)^2\right| \\
=\; &\left|\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)} \left(\widehat{\varepsilon}_{\underline{t}}^2 - \varepsilon_{\underline{t}}^2\right) - 2\cdot\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)} \varepsilon_{\underline{t}}\left(\widehat{\varepsilon}_{\underline{t}} - \varepsilon_{\underline{t}}\right)\right| \\
\le\; &\left|\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)} \left(\widehat{\varepsilon}_{\underline{t}}^2 - \varepsilon_{\underline{t}}^2\right)\right| + \mathcal{O}_P(1)\cdot\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)} \left|Q_{\underline{t}} + R_{\underline{t}} - \overline{\varepsilon}\right| \\
=\; &o_P(1),
\end{aligned} \tag{2.90}$$

from (2.82) (with $w = 1$), (2.87), (2.89) and $(IV) = o_P(1)$. □

Proof of Lemma 2.13, assertion (2.21):
This assertion can be obtained following exactly along the lines of the proof of Lemma 5.5 and Corollary 5.6 of Bühlmann (1997). The only difference to Bühlmann's proof is that we decompose

$$X^*_{\underline{t}} = X^*_{\underline{t},M} + U^*_{\underline{t}} + V^*_{\underline{t}},$$

with $X^*_{\underline{t},M}$ as defined in (2.18) and

$$U^*_{\underline{t}} := \sum_{\underline{k}\in\Theta(M)} \left(\widehat{b}_{\underline{k}}(p) - b_{\underline{k}}\right)\varepsilon^*_{\underline{t}-\underline{k}}, \quad \text{and} \quad V^*_{\underline{t}} := \sum_{\underline{k}\in\Theta\setminus\Theta(M)} \widehat{b}_{\underline{k}}(p)\,\varepsilon^*_{\underline{t}-\underline{k}},$$

analogously for $\widetilde{X}_{\underline{t}}$. The only assertions needed to adapt the proof of Bühlmann's Lemma 5.5 are given by (2.19) and (2.20), which correspond to Lemmas 5.1 and 5.3 in Bühlmann (1997), as well as

$$\sum_{\underline{k}\in\Theta} \left|\widehat{b}_{\underline{k}}(p) - b_{\underline{k}}\right| = o_P(1) \tag{2.91}$$

and

$$\varepsilon^*_{\underline{t}} \xrightarrow{d^*} \varepsilon_{\underline{t}} \quad \text{in prob.}, \tag{2.92}$$

which correspond to Lemmas 5.2 and 5.4 in Bühlmann (1997). In the following, we complete the proof by showing the latter two assertions.
It holds

$$\begin{aligned}
&\sum_{\underline{k}\in\Theta} \left|\widehat{b}_{\underline{k}}(p) - b_{\underline{k}}\right| \\
\leq\ & \sum_{\underline{k}\in\Theta} \left|\widehat{b}_{\underline{k}}(p) - b_{\underline{k}}(p)\right| + \sum_{\underline{k}\in\Theta} \left|b_{\underline{k}}(p) - b_{\underline{k}}\right| \\
\leq\ & \frac{1}{p^4}\, \mathcal{O}_P(1) \cdot \sum_{k_1=-\infty}^{\infty} \sum_{k_2=0}^{\infty} \left(\frac{p}{p+1}\right)^{|k_1|+k_2} + C \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1 + |\underline{k}|_\infty)\, |a_{\underline{k}}| \\
\leq\ & \mathcal{O}_P(1) \cdot \frac{2(p+1)^2}{p^4} + C \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} (1 + |\underline{k}|_\infty)\, |a_{\underline{k}}| \\
=\ & o_P(1),
\end{aligned}$$

due to Lemma 2.9, Lemma 2.10, (2.6) and Assumption 2. This yields (2.91).
As for (2.92), we adapt the proof of Lemma 5.4 of Bühlmann (1997). Let F with $F(x) := P\{\varepsilon_{\underline{t}} \leq x\} = P\{\widetilde{\varepsilon}_{\underline{t}} \leq x\}$ be the distribution function of $(\varepsilon_{\underline{t}})$ and let F_n be the empirical distribution function of $\{\varepsilon_{\underline{t}} : \underline{t} \in \Pi(n,p)\}$ as defined in Assumption 4. Furthermore, according to step 2 of the AR sieve bootstrap procedure, $(\varepsilon^*_{\underline{t}})$ is an i.i.d. sequence with marginal distribution function $\widehat{F}_n$, where

$$\widehat{F}_n(x) = \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \mathbb{1}\{\widehat{\varepsilon}_{\underline{t}}(p) \leq x\},$$

and $\widehat{\varepsilon}_{\underline{t}}(p) = \varepsilon'_{\underline{t}}(p) - \overline{\varepsilon}$ are the centered residuals of the autoregressive fit with $\varepsilon'_{\underline{t}}(p) = X_{\underline{t}} - \sum_{\underline{k}\in\Theta(p)} \widehat{a}_{\underline{k}}(p)\, X_{\underline{t}-\underline{k}}$ and $\overline{\varepsilon} = (1/|\Pi(n,p)|) \sum_{\underline{t}\in\Pi(n,p)} \varepsilon'_{\underline{t}}(p)$. We use the Mallows metric d_2, cf. Bickel and Freedman (1981), and derive

$$d_2(\widehat{F}_n, F) \leq d_2(\widehat{F}_n, F_n) + d_2(F_n, F).$$

From Assumption 4 we have convergence of second moments

$$\left|\int x^2\, dF_n(x) - \int x^2\, dF(x)\right| = \left|\frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} (\varepsilon_{\underline{t}})^2 - E(\varepsilon_{\underline{0}})^2\right| = o_P(1).$$

This, together with Assumption 4, implies $d_2(F_n, F) = o_P(1)$, according to Lemma 8.3 of Bickel and Freedman (1981). Now let S be uniformly distributed on the finite set $\Pi(n,p)$. For any given realizations of $\{\varepsilon_{\underline{t}} : \underline{t} \in \Pi(n,p)\}$ and $\{\widehat{\varepsilon}_{\underline{t}}(p) : \underline{t} \in \Pi(n,p)\}$, $\widehat{F}_n$ and F_n are deterministic distribution functions, and it is easy to see that ε_S has distribution function F_n and $\widehat{\varepsilon}_S(p)$ has distribution function $\widehat{F}_n$. Hence, it holds

$$d_2(\widehat{F}_n, F_n) \leq E_S\Big(\widehat{\varepsilon}_S(p) - \varepsilon_S\Big)^2 \leq \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \Big(\widehat{\varepsilon}_{\underline{t}}(p) - \varepsilon_{\underline{t}}\Big)^2.$$

Therefore, we have for the random variable $d_2(\widehat{F}_n, F_n)$:

$$d_2(\widehat{F}_n, F_n) \leq \frac{1}{|\Pi(n,p)|} \sum_{\underline{t}\in\Pi(n,p)} \Big(\widehat{\varepsilon}_{\underline{t}}(p) - \varepsilon_{\underline{t}}\Big)^2 = o_P(1),$$

due to (2.90). This implies $d_2(\widehat{F}_n, F) = o_P(1)$, and, therefore, (2.92). □

Proof of Lemma 2.13, assertion (2.22):

Let $\underline{c} \in \mathbb{R}^k$ be arbitrary. Using the notation $\|z\|_q = \Big(E(|z|^q)\Big)^{1/q}$, the goal is to prove $\|\underline{c}^T\, g(\widetilde{\mathbf{Y}}_{\underline{t},M})\|_{2+2/(h+1)} \leq C$ uniformly for all $\underline{t} \in \mathbb{Z}^2$. Due to

$$\Big\|\underline{c}^T\, g(\widetilde{\mathbf{Y}}_{\underline{t},M})\Big\|_{2+2/(h+1)} \leq \sum_{v=1}^{k} |c_v|\, \Big\|g_v(\widetilde{\mathbf{Y}}_{\underline{t},M})\Big\|_{2+2/(h+1)}$$

it suffices to show $\|g_v(\widetilde{\mathbf{Y}}_{\underline{t},M})\|_{2+2/(h+1)} \leq C_v$ for all $v = 1, \ldots, k$ (note that C may depend on $\underline{c}$). We will derive this assertion from a slight modification of Lemma 2.14. One can easily observe that the assertion of Lemma 2.14 remains true if one replaces $\widetilde{\mathbf{Y}}_{\underline{t}}$ with $\widetilde{\mathbf{Y}}_{\underline{t},M}$, i.e. it holds for each $W \subset \Theta \cup \{\underline{0}\}$

$$\Big\|g_v(\widetilde{\mathbf{Y}}_{\underline{t},M}) - g_v(\widetilde{\mathbf{Y}}^{(W)}_{\underline{t},M})\Big\|_2 \;\leq\; C_v \cdot \left(\sum_{\underline{k}\in\Theta\setminus W} |b_{\underline{k}}| + \mathbb{1}_{\{\underline{0}\notin W\}}\right).$$

Now we modify the proof of Lemma 2.14 by choosing $W = \emptyset$, which yields $\widetilde{\mathbf{Y}}^{(W)}_{\underline{t},M} = \underline{0}$, and by replacing the $\|\cdot\|_2$-norm with $\|\cdot\|_{2+2/(h+1)}$. Then, (2.99) reads:

$$\left\|\frac{D^{\boldsymbol{\alpha}} g_v(\underline{0})}{\boldsymbol{\alpha}!}\,(\widetilde{\mathbf{Y}}_{\underline{t},M})^{\boldsymbol{\alpha}}\right\|_{2+2/(h+1)} \leq C \cdot \left\|\frac{D^{\boldsymbol{\alpha}} g_v(\underline{0})}{\boldsymbol{\alpha}!}\right\|_{\left(2+\frac{2}{h+1}\right)\left(\frac{h+2}{h+2-|\boldsymbol{\alpha}|}\right)} \cdot \left(\sum_{\underline{k}\in\Theta} |b_{\underline{k}}| + 1\right)^{|\boldsymbol{\alpha}|}.$$

Note that the expression on the right-hand side does not depend on $\underline{t}$ and that

$$\left\|\frac{D^{\boldsymbol{\alpha}} g_v(\underline{0})}{\boldsymbol{\alpha}!}\right\|_{\left(2+\frac{2}{h+1}\right)\left(\frac{h+2}{h+2-|\boldsymbol{\alpha}|}\right)} = \left|\frac{D^{\boldsymbol{\alpha}} g_v(\underline{0})}{\boldsymbol{\alpha}!}\right| < \infty,$$

because the derivative of g_v at the origin is deterministic. Along the lines of the proof of Lemma 2.14, with the modifications mentioned above, one obtains

$$\left\| g_v(\widetilde{\mathbf{Y}}_{\underline{t},M}) - g_v(\underline{0}) \right\|_{2+2/(h+1)} \leq C_v \cdot \left(\sum_{\underline{k} \in \Theta} |b_{\underline{k}}| + 1 \right),$$

which completes the proof of the second assertion of (2.22) via

$$\begin{aligned} \left\| g_v(\widetilde{\mathbf{Y}}_{\underline{t},M}) \right\|_{2+2/(h+1)} &\leq \left\| g_v(\underline{0}) \right\|_{2+2/(h+1)} + \left\| g_v(\widetilde{\mathbf{Y}}_{\underline{t},M}) - g_v(\underline{0}) \right\|_{2+2/(h+1)} \\ &\leq \left| g_v(\underline{0}) \right| + C_v \cdot \left(\sum_{\underline{k} \in \Theta} |b_{\underline{k}}| + 1 \right). \end{aligned}$$

An analogous modification for the bootstrap quantities in Lemma 2.14 yields

$$E^*\Big(\left| \underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) \right|^{2+2/(h+1)} \Big) = \Big(\left\| \underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}) \right\|_{2+2/(h+1)} \Big)^{2+2/(h+1)} = \mathcal{O}_P(1),$$

with exactly the same arguments as for the non-bootstrap quantities. □

Proof of Lemma 2.13, assertion (2.23):
For arbitrary but fixed $\underline{c} \in \mathbb{R}^k$ we abbreviate $l(\underline{x}) := \underline{c}^T g(\underline{x})$. Let $0 < K < \infty$ be a constant that will be specified later on. We define the K-truncated version of function l by

$$\widetilde{l}(\underline{x}) := l(\underline{x}) \cdot \mathbb{1}\{|l(\underline{x})| \leq K\} + K \cdot \mathrm{sgn}(l(\underline{x})) \cdot \mathbb{1}\{|l(\underline{x})| > K\}.$$

For arbitrary $\varepsilon > 0$ we get from standard calculations

$$\begin{aligned} & P\Big\{ \Big| \mathrm{Cov}^*\Big(l(\mathbf{Y}^*_{\underline{h},M}), l(\mathbf{Y}^*_{\underline{0},M}) \Big) - \mathrm{Cov}\Big(l(\widetilde{\mathbf{Y}}_{\underline{h},M}), l(\widetilde{\mathbf{Y}}_{\underline{0},M}) \Big) \Big| > \varepsilon \Big\} \\ \leq\ & P\{\, |I| > \varepsilon/3 \,\} + P\{\, |II| > \varepsilon/3 \,\} + P\{\, |III| > \varepsilon/3 \,\}, \end{aligned} \tag{2.93}$$

where

$$\begin{aligned} I &:= \mathrm{Cov}^*\Big(l(\mathbf{Y}^*_{\underline{h},M}), l(\mathbf{Y}^*_{\underline{0},M}) \Big) - \mathrm{Cov}^*\Big(\widetilde{l}(\mathbf{Y}^*_{\underline{h},M}), \widetilde{l}(\mathbf{Y}^*_{\underline{0},M}) \Big), \\ II &:= \mathrm{Cov}^*\Big(\widetilde{l}(\mathbf{Y}^*_{\underline{h},M}), \widetilde{l}(\mathbf{Y}^*_{\underline{0},M}) \Big) - \mathrm{Cov}\Big(\widetilde{l}(\widetilde{\mathbf{Y}}_{\underline{h},M}), \widetilde{l}(\widetilde{\mathbf{Y}}_{\underline{0},M}) \Big), \\ III &:= \mathrm{Cov}\Big(\widetilde{l}(\widetilde{\mathbf{Y}}_{\underline{h},M}), \widetilde{l}(\widetilde{\mathbf{Y}}_{\underline{0},M}) \Big) - \mathrm{Cov}\Big(l(\widetilde{\mathbf{Y}}_{\underline{h},M}), l(\widetilde{\mathbf{Y}}_{\underline{0},M}) \Big). \end{aligned}$$

Hence, the desired assertion follows if we can, for each $\delta > 0$, specify $0 < K < \infty$ and $n_0 \in \mathbb{N}$ such that the right-hand side of (2.93) is smaller than δ for all $n \geq n_0$.

Per definition, $l(\underline{x})$ can be expanded as

$$l(\underline{x}) = \widetilde{l}(\underline{x}) + \Big[l(\underline{x}) - K \cdot \mathrm{sgn}(l(\underline{x}))\Big] \cdot \mathbb{1}\{|l(\underline{x})| > K\}.$$

Using this, we get

$$\begin{aligned} I \;=\; & \mathrm{Cov}^*\Big(\widetilde{l}(\mathbf{Y}^*_{\underline{h},M}), \Big[l(\mathbf{Y}^*_{\underline{0},M}) - K \cdot \mathrm{sgn}(l(\mathbf{Y}^*_{\underline{0},M}))\Big] \cdot \mathbb{1}\{|l(\mathbf{Y}^*_{\underline{0},M})| > K\}\Big) \\ & + \mathrm{Cov}^*\Big(\Big[l(\mathbf{Y}^*_{\underline{h},M}) - K \cdot \mathrm{sgn}(l(\mathbf{Y}^*_{\underline{h},M}))\Big] \cdot \mathbb{1}\{|l(\mathbf{Y}^*_{\underline{h},M})| > K\}, \widetilde{l}(\mathbf{Y}^*_{\underline{0},M})\Big) \\ & + \mathrm{Cov}^*\Big(\Big[l(\mathbf{Y}^*_{\underline{h},M}) - K \cdot \mathrm{sgn}(l(\mathbf{Y}^*_{\underline{h},M}))\Big] \cdot \mathbb{1}\{|l(\mathbf{Y}^*_{\underline{h},M})| > K\}, \\ & \qquad \Big[l(\mathbf{Y}^*_{\underline{0},M}) - K \cdot \mathrm{sgn}(l(\mathbf{Y}^*_{\underline{0},M}))\Big] \cdot \mathbb{1}\{|l(\mathbf{Y}^*_{\underline{0},M})| > K\}\Big). \end{aligned}$$

The first summand on the right-hand side can be bounded in absolute value with Hölder's and Markov's inequalities by

$$\begin{aligned} & \Big(E^*\Big(\widetilde{l}(\mathbf{Y}^*_{\underline{h},M})^2\Big)\Big)^{1/2} \cdot \Big(E^*\Big(\Big|l(\mathbf{Y}^*_{\underline{0},M})\Big|^2 \cdot \mathbb{1}\{|l(\mathbf{Y}^*_{\underline{0},M})| > K\}\Big)\Big)^{1/2} \\ \leq\; & \mathcal{O}_P(1) \cdot \Big(E^*\Big(\Big|l(\mathbf{Y}^*_{\underline{0},M})\Big|^{2(h+2)/(h+1)}\Big)\Big)^{(h+1)/2(h+2)} \cdot \Big(P^*\Big\{|l(\mathbf{Y}^*_{\underline{0},M})| > K\Big\}\Big)^{1/2(h+2)} \\ \leq\; & \mathcal{O}_P(1) \cdot \Big(E^*\Big(\Big|l(\mathbf{Y}^*_{\underline{0},M})\Big|^{2(h+2)/(h+1)}\Big)\Big)^{1/2} \cdot \Big(\frac{1}{K^{2(h+2)/(h+1)}}\Big)^{1/2(h+2)} \\ =\; & K^{-1/(h+1)} \cdot \mathcal{O}_P(1), \end{aligned}$$

where the boundedness in probability of the moments is taken from (2.22), noting that $2(h+2)/(h+1) = 2 + 2/(h+1)$. The same calculations can be done for the second and third summand above which yields $I = K^{-1/(h+1)} \cdot \mathcal{O}_P(1)$. Hence, for the given $\delta > 0$, there exists $S(\delta) < \infty$ such that

$$P\Big\{|I| > S(\delta)/K^{1/(h+1)}\Big\} \leq \delta/2 \quad \forall\, n \in \mathbb{N},$$

and for each $K > (3\, S(\delta)/\varepsilon)^{h+1}$ we have

$$P\Big\{|I| > \varepsilon/3\Big\} \leq P\Big\{|I| > S(\delta)/K^{1/(h+1)}\Big\} \leq \delta/2 \quad \forall\, n \in \mathbb{N}.$$

By the very same calculations as for I, replacing E^* with E, one obtains

$$|III| \leq \widetilde{C} \cdot K^{-1/(h+1)} \quad \forall\, n \in \mathbb{N}$$

for some $\widetilde{C} < \infty$, using $E\,|l(\widetilde{\mathbf{Y}}_{\underline{t},M})|^{(2+2/(h+1))} \leq C$, cf. (2.22). Choosing $K > (3\,\widetilde{C}/\varepsilon)^{h+1}$ gives

$$P\Big\{|III| > \varepsilon/3\Big\} \leq P\Big\{|III| > \widetilde{C}/K^{1/(h+1)}\Big\} = 0 \quad \forall\, n \in \mathbb{N},$$

noting that III is deterministic. Combining the results for I and III, we get from choosing $K > (3\,(\widetilde{C} \vee S(\delta))/\varepsilon)^{h+1}$

$$P\Big\{|I| > \varepsilon/3\Big\} + P\Big\{|III| > \varepsilon/3\Big\} \le \delta/2 \quad \forall\, n \in \mathbb{N}. \tag{2.94}$$

For this fixed $K < \infty$ we will now show $II = o_P(1)$. (2.21) implies

$$\begin{pmatrix} \mathbf{Y}^*_{\underline{h},M} \\ \mathbf{Y}^*_{\underline{0},M} \end{pmatrix} \xrightarrow{d^*} \begin{pmatrix} \widetilde{\mathbf{Y}}_{\underline{h},M} \\ \widetilde{\mathbf{Y}}_{\underline{0},M} \end{pmatrix} \quad \text{in } P\text{-prob.}$$

Hence, we have $E^* f(\mathbf{Y}^*_{\underline{h},M}, \mathbf{Y}^*_{\underline{0},M}) \to E\, f(\widetilde{\mathbf{Y}}_{\underline{h},M}, \widetilde{\mathbf{Y}}_{\underline{0},M})$ in P-probability for each continuous and bounded function f. It follows

$$\begin{aligned} II &= E^*\Big(\widetilde{l}(\mathbf{Y}^*_{\underline{h},M})\,\widetilde{l}(\mathbf{Y}^*_{\underline{0},M})\Big) - E\Big(\widetilde{l}(\widetilde{\mathbf{Y}}_{\underline{h},M})\,\widetilde{l}(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big) \\ &\quad + E^*\Big(\widetilde{l}(\mathbf{Y}^*_{\underline{h},M})\Big)\, E^*\Big(\widetilde{l}(\mathbf{Y}^*_{\underline{0},M})\Big) - E\Big(\widetilde{l}(\widetilde{\mathbf{Y}}_{\underline{h},M})\Big)\, E\Big(\widetilde{l}(\widetilde{\mathbf{Y}}_{\underline{0},M})\Big) \\ &= o_P(1), \end{aligned}$$

since $\widetilde{l}$ is continuous and bounded by K. We can therefore find $n_0 \in \mathbb{N}$ such that

$$P\{\,|II| > \varepsilon/3\,\} \le \delta/2 \quad \forall\, n \ge n_0,$$

which together with (2.93) and (2.94) completes the proof. □

Proof of Lemma 2.13, assertion (2.24):
We prove that $\sum_{|h_1|,|h_2| \le M} \Big|\mathrm{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\Big|$ converges to a finite limit as $M \to \infty$, by showing that the series tails

$$\sum_{h_1=M+1}^{\infty} \sum_{h_2=-M}^{M} \Big|\mathrm{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\Big| + \sum_{h_1=-\infty}^{-M-1} \sum_{h_2=-M}^{M} \Big|\mathrm{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\Big| \tag{2.95}$$

as well as the remaining tails

$$\sum_{h_1=-M}^{M} \sum_{|h_2| \ge M+1} \Big|\ldots\Big|, \qquad \sum_{|h_1|,|h_2|=M+1}^{\infty} \Big|\ldots\Big| \tag{2.96}$$

vanish for $M \to \infty$. We will only consider the first summand in (2.95) because all other expressions, including the ones in (2.96), can be treated with analogous arguments. In accordance with the definition of the vector $\widetilde{\mathbf{Y}}_{\underline{t}}$ we define for each $\underline{h} \in \mathbb{Z}^2$ the truncated versions $\widetilde{\mathbf{Y}}^{(l)}_{\underline{h}}$ (truncated at the left-hand side) and $\widetilde{\mathbf{Y}}^{(r)}_{\underline{0}}$ (truncated at the right-hand side) via

$$\widetilde{\mathbf{Y}}^{(l)}_{\underline{h}} := (\widetilde{X}^{(l)}_{\underline{h}+\underline{s}(1)}, \ldots, \widetilde{X}^{(l)}_{\underline{h}+\underline{s}(m_1 m_2)})^T, \quad \widetilde{\mathbf{Y}}^{(r)}_{\underline{0}} := (\widetilde{X}^{(r)}_{\underline{0}+\underline{s}(1)}, \ldots, \widetilde{X}^{(r)}_{\underline{0}+\underline{s}(m_1 m_2)})^T,$$

where

$$
\begin{aligned}
\widetilde{X}^{(l)}_{\underline{h}+\underline{s}(j)} &:= \sum_{\underline{k}\in\Theta} b_{\underline{k}} \cdot \mathbb{1}\{k_1 \leq \lfloor h_1/2 \rfloor - m_1\}\, \widetilde{\varepsilon}_{\underline{h}+\underline{s}(j)-\underline{k}} + \widetilde{\varepsilon}_{\underline{h}+\underline{s}(j)},\\
\widetilde{X}^{(r)}_{\underline{0}+\underline{s}(j)} &:= \sum_{\underline{k}\in\Theta} b_{\underline{k}} \cdot \mathbb{1}\{k_1 \geq -\lfloor h_1/2 \rfloor\}\, \widetilde{\varepsilon}_{\underline{0}+\underline{s}(j)-\underline{k}} + \widetilde{\varepsilon}_{\underline{0}+\underline{s}(j)}.
\end{aligned}
$$

Note that the dependence of $\widetilde{\mathbf{Y}}^{(r)}_{\underline{0}}$ on $\underline{h}$ is suppressed in the notation. One can easily check that $\widetilde{\mathbf{Y}}^{(l)}_{\underline{h}}$ and $\widetilde{\mathbf{Y}}^{(r)}_{\underline{0}}$ are independent random variables. Hence, the first summand in (2.95) can be bounded by

$$
\begin{aligned}
&\sum_{h_1=M+1}^{\infty} \sum_{h_2=-M}^{M} \Big|\mathrm{Cov}\Big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}) - g_u(\widetilde{\mathbf{Y}}^{(l)}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\Big)\Big|\\
+\;&\sum_{h_1=M+1}^{\infty} \sum_{h_2=-M}^{M} \Big|\mathrm{Cov}\Big(g_u(\widetilde{\mathbf{Y}}^{(l)}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}}) - g_v(\widetilde{\mathbf{Y}}^{(r)}_{\underline{0}})\Big)\Big|.
\end{aligned}
$$

Both of these expressions can be treated in the same way. Therefore, we will only consider the latter summand which can be bounded by

$$
C \cdot \sum_{h_1=M+1}^{\infty} \sum_{h_2=-M}^{M} \Big\|g_v(\widetilde{\mathbf{Y}}_{\underline{0}}) - g_v(\widetilde{\mathbf{Y}}^{(r)}_{\underline{0}})\Big\|_2,
$$

because $\|g_u(\widetilde{\mathbf{Y}}^{(l)}_{\underline{h}})\|_2 \leq C$ follows as in (2.22) (here, $\|\cdot\|_w$ denotes the usual L^w-norm). For the remainder of this proof, C will denote a generic constant that may change from line to line. We get from Lemma 2.14

$$
\Big\|g_v(\widetilde{\mathbf{Y}}_{\underline{0}}) - g_v(\widetilde{\mathbf{Y}}^{(r)}_{\underline{0}})\Big\|_2 \leq C \cdot \sum_{k_1=-\infty}^{-\lfloor h_1/2 \rfloor -1} \sum_{k_2=1}^{\infty} |b_{\underline{k}}|. \tag{2.97}
$$

Hence, the previously derived expression is bounded by

$$
\begin{aligned}
&C \cdot \sum_{h_1=M+1}^{\infty} \sum_{h_2=-M}^{M} \sum_{k_1=-\infty}^{-\lfloor h_1/2 \rfloor -1} \sum_{k_2=1}^{\infty} |b_{\underline{k}}|\\
\leq\;& C \cdot (2M+1) \sum_{k_1=-\infty}^{-\lfloor (M+1)/2 \rfloor -1} 2\Big(- \lfloor (M+1)/2 \rfloor - k_1\Big) \sum_{k_2=1}^{\infty} |b_{\underline{k}}|\\
\leq\;& C \cdot \sum_{k_1=-\infty}^{-\lfloor (M+1)/2 \rfloor -1} \sum_{k_2=1}^{\infty} M\,|k_1|\,|b_{\underline{k}}| \leq C \cdot \sum_{k_1=-\infty}^{-\lfloor (M+1)/2 \rfloor -1} \sum_{k_2=1}^{\infty} (1+|\underline{k}|_\infty)^2\,|b_{\underline{k}}|,
\end{aligned}
$$

since it holds $M \leq 2\,|k_1| \leq 2\,|\underline{k}|_\infty$ for all $k_1 \leq -\lfloor (M+1)/2 \rfloor - 1$. Note that the right-hand side converges to zero as $M \to \infty$, because Lemma 2.1 ensures

$\sum_{\underline{k}\in\Theta}(1+|\underline{k}|_\infty)^{r-1}\,|b_{\underline{k}}| < \infty$, and we assume $r=4$ in Lemma 2.13. The remaining expressions in (2.95) and (2.96) can be treated analogously, using the summability conditions $\sum_{\underline{k}\in\Theta}|k_2|^2\,|b_{\underline{k}}| < \infty$ and $\sum_{\underline{k}\in\Theta}|k_1\,k_2|\,|b_{\underline{k}}| < \infty$ which are fulfilled since $|k_2|^2 \le (1+|\underline{k}|_\infty)^2$ and $|k_1\,k_2| \le (1+|\underline{k}|_\infty)^2$. This completes the proof. $\square$

Proof of Lemma 2.14:

We will perform a Taylor expansion of order h of $g_v(\widetilde{\mathbf{Y}}_{\underline{t}})$ around $\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}$. Let $m := m_1 m_2$. We use the common multi-index notation $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_m) \in \mathbb{N}_0^m$ with $|\boldsymbol{\alpha}| = \sum_{i=1}^m \alpha_i$ and $\boldsymbol{\alpha}! = \alpha_1!\cdot\ldots\cdot\alpha_m!$. Furthermore, we abbreviate

$$D^{\boldsymbol{\alpha}} g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}) := \left.\frac{\partial^{|\boldsymbol{\alpha}|} g_v(\underline{x})}{\partial {x_1}^{\alpha_1}\ldots\partial {x_m}^{\alpha_m}}\right|_{\underline{x}=\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}},$$

and get

$$\begin{aligned}
&\left\| g_v(\widetilde{\mathbf{Y}}_{\underline{t}}) - g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})\right\|_2 \\
&\le \sum_{1\le|\boldsymbol{\alpha}|<h}\left\|\frac{D^{\boldsymbol{\alpha}} g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})}{\boldsymbol{\alpha}!}\,(\widetilde{\mathbf{Y}}_{\underline{t}} - \widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\alpha}}\right\|_2 + \sum_{|\boldsymbol{\alpha}|=h}\left\|\frac{D^{\boldsymbol{\alpha}} g_v(\boldsymbol{\xi}_{\underline{t}})}{\boldsymbol{\alpha}!}\,(\widetilde{\mathbf{Y}}_{\underline{t}} - \widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\alpha}}\right\|_2, \qquad (2.98)
\end{aligned}$$

where $\boldsymbol{\xi}_{\underline{t}}$ is between $\widetilde{\mathbf{Y}}_{\underline{t}}$ and $\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}$. Note that for each $\boldsymbol{\alpha}$ we find suitable integers $1 \le j(1), j(2), \ldots, j(|\boldsymbol{\alpha}|) \le m$ such that

$$(\widetilde{\mathbf{Y}}_{\underline{t}} - \widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\alpha}} = (\widetilde{X}_{\underline{t}+\underline{s}(j(1))} - \widetilde{X}_{\underline{t}+\underline{s}(j(1))}^{(W)})\cdot\ldots\cdot(\widetilde{X}_{\underline{t}+\underline{s}(j(|\boldsymbol{\alpha}|))} - \widetilde{X}_{\underline{t}+\underline{s}(j(|\boldsymbol{\alpha}|))}^{(W)})$$

and, thus, Hölder's inequality yields

$$\begin{aligned}
&\left\|\frac{D^{\boldsymbol{\alpha}} g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})}{\boldsymbol{\alpha}!}\,(\widetilde{\mathbf{Y}}_{\underline{t}} - \widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\alpha}}\right\|_2 \\
\le\ & \left\|\frac{D^{\boldsymbol{\alpha}} g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})}{\boldsymbol{\alpha}!}\right\|_{2(h+2)/(h+2-|\boldsymbol{\alpha}|)} \cdot \prod_{k=1}^{|\boldsymbol{\alpha}|}\left\|\widetilde{X}_{\underline{t}+\underline{s}(j(k))} - \widetilde{X}_{\underline{t}+\underline{s}(j(k))}^{(W)}\right\|_{2(h+2)} \\
\le\ & C\cdot\left\|\frac{D^{\boldsymbol{\alpha}} g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})}{\boldsymbol{\alpha}!}\right\|_{2(h+2)/(h+2-|\boldsymbol{\alpha}|)} \cdot \left(\sum_{\underline{k}\in\Theta\setminus W}|b_{\underline{k}}| + \mathbb{1}_{\{\underline{0}\notin W\}}\right)^{|\boldsymbol{\alpha}|}. \qquad (2.99)
\end{aligned}$$

Here, we have used that it follows, per definition, for any index $\underline{u}$

$$\begin{aligned}
\left\|\widetilde{X}_{\underline{u}} - \widetilde{X}_{\underline{u}}^{(W)}\right\|_{2(h+2)} &= \left\|\sum_{\underline{k}\in\Theta\setminus W} b_{\underline{k}}\,\widetilde{\varepsilon}_{\underline{u}-\underline{k}} + \widetilde{\varepsilon}_{\underline{u}}\,\mathbb{1}_{\{\underline{0}\notin W\}}\right\|_{2(h+2)} \\
&\le \|\widetilde{\varepsilon}_{\underline{0}}\|_{2(h+2)}\left(\sum_{\underline{k}\in\Theta\setminus W}|b_{\underline{k}}| + \mathbb{1}_{\{\underline{0}\notin W\}}\right) \qquad (2.100)
\end{aligned}$$

from strict stationarity of $(\widetilde{\varepsilon}_{\underline{t}})$. Note that $\|\widetilde{\varepsilon}_{\underline{0}}\|_{2(h+2)} < \infty$ follows from Assumption 4. On the other hand, abbreviating $q := 2(h+2)/(h+2-|\boldsymbol{\alpha}|)$, the first factor in (2.99) can be bounded via another Taylor expansion of order $h - |\boldsymbol{\alpha}|$ around the zero vector as

$$
\begin{aligned}
&\left\| \frac{D^{\boldsymbol{\alpha}} g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})}{\boldsymbol{\alpha}!} \right\|_q \\
\leq \quad & \sum_{0 \leq |\boldsymbol{\beta}| < h - |\boldsymbol{\alpha}|} \left\| \frac{D^{\boldsymbol{\alpha}+\boldsymbol{\beta}} g_v(\underline{0})}{\boldsymbol{\alpha}!\,\boldsymbol{\beta}!} \,(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\beta}} \right\|_q + \sum_{|\boldsymbol{\beta}| = h - |\boldsymbol{\alpha}|} \left\| \frac{D^{\boldsymbol{\alpha}+\boldsymbol{\beta}} g_v(\boldsymbol{\tau}_{\underline{0}})}{\boldsymbol{\alpha}!\,\boldsymbol{\beta}!} \,(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\beta}} \right\|_q ,
\end{aligned}
$$

where $\boldsymbol{\tau}_{\underline{0}}$ is between $\underline{0}$ and $\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}$. The first summand on the right-hand side, analogous to (2.99) and (2.100), is bounded by

$$
C \cdot \sum_{|\boldsymbol{\beta}| < h - |\boldsymbol{\alpha}|} \left(1 + \sum_{\underline{k} \in \Theta} |b_{\underline{k}}| \right)^{|\boldsymbol{\beta}|} < \infty,
$$

since the derivative at zero is constant and $q|\boldsymbol{\beta}| \leq 2(h+2)$. Using the Lipschitz property of the h-th derivatives of g_v, the second summand is bounded by

$$
\begin{aligned}
& \sum_{|\boldsymbol{\beta}| = h - |\boldsymbol{\alpha}|} \left(\left\| \frac{D^{\boldsymbol{\alpha}+\boldsymbol{\beta}} g_v(\underline{0})}{\boldsymbol{\alpha}!\,\boldsymbol{\beta}!} \,(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\beta}} \right\|_q + \left\| \left| \frac{D^{\boldsymbol{\alpha}+\boldsymbol{\beta}} g_v(\boldsymbol{\tau}_{\underline{0}})}{\boldsymbol{\alpha}!\,\boldsymbol{\beta}!} - \frac{D^{\boldsymbol{\alpha}+\boldsymbol{\beta}} g_v(\underline{0})}{\boldsymbol{\alpha}!\,\boldsymbol{\beta}!} \right| (\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\beta}} \right\|_q \right) \\
\leq \quad & \sum_{|\boldsymbol{\beta}| = h - |\boldsymbol{\alpha}|} \left(\left\| \frac{D^{\boldsymbol{\alpha}+\boldsymbol{\beta}} g_v(\underline{0})}{\boldsymbol{\alpha}!\,\boldsymbol{\beta}!} \,(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\beta}} \right\|_q + C \cdot \left\| \left(\sum_{j=1}^{m} \left| \widetilde{X}_{\underline{t}+\underline{s}(j)}^{(W)} \right| \right) (\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)})^{\boldsymbol{\beta}} \right\|_q \right),
\end{aligned}
$$

which is finite due to similar arguments as for the first summand. With the same calculation we can also treat the second sum in (2.98) analogous to the first sum. Together with (2.99) and (2.100), we finally get

$$
\begin{aligned}
\left\| g_v(\widetilde{\mathbf{Y}}_{\underline{t}}) - g_v(\widetilde{\mathbf{Y}}_{\underline{t}}^{(W)}) \right\|_2 \;&\leq\; C \cdot \sum_{1 \leq |\boldsymbol{\alpha}| \leq h} \left(\sum_{\underline{k} \in \Theta \setminus W} |b_{\underline{k}}| + \mathbb{1}_{\{\underline{0} \notin W\}} \right)^{|\boldsymbol{\alpha}|} \\
&\leq\; C \cdot \left(\sum_{\underline{k} \in \Theta \setminus W} |b_{\underline{k}}| + \mathbb{1}_{\{\underline{0} \notin W\}} \right), \qquad (2.101)
\end{aligned}
$$

with a generic constant $C < \infty$, which depends merely on $\|\widetilde{\varepsilon}_{\underline{0}}\|_{2(h+2)}$ and on $\sum_{\underline{k} \in \Theta} |b_{\underline{k}}|$. Therefore, one can follow along these lines for the second assertion in Lemma 2.14 concerning the bootstrap versions $\mathbf{Y}_{\underline{t}}^{*}$ and $\mathbf{Y}_{\underline{t}}^{*(W)}$. Since (2.20), Assumption 4 and (2.19) ensure $\|\varepsilon_{\underline{t}}^{*}\|_{*2(h+2)} = \mathcal{O}_P(1)$ and $\sum_{\underline{k} \in \Theta} \left| \widehat{b}_{\underline{k}}(p) \right| = \mathcal{O}_P(1)$, it follows with the same calculation as for (2.101)

$$\left\| g_v(\mathbf{Y}_{\underline{t}}^*) - g_v(\mathbf{Y}_{\underline{t}}^{*(W)}) \right\|_{*2} \;\leq\; \mathcal{O}_P(1) \cdot \left(\sum_{\underline{k} \in \Theta \setminus W} \left| \widehat{b}_{\underline{k}}(p) \right| + \mathbb{1}_{\{\underline{0} \notin W\}} \right),$$

which completes the proof. □

Proof of Lemma 2.16:

Suppose $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ is a linear spatial process as given by (2.27) with some suitable absolutely summable coefficients $(\alpha_{\underline{\nu}})_{\underline{\nu}\in\mathbb{Z}^2}$ and an i.i.d. white noise process $(u_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ with $E(u_{\underline{t}}^2) = \sigma^2 \in (0,\infty)$ and $E(u_{\underline{t}}^4) = \eta\sigma^4 \in (0,\infty)$. For the comparative quantity $\check{\gamma}(\underline{h})$ defined in Lemma 2.16 it holds

$$\check{\gamma}(\underline{h}) := \frac{1}{n^2} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} X_{\underline{t}+\underline{h}}\, X_{\underline{t}},$$

which is asymptotically equivalent to $\widehat{\gamma}(\underline{h})$. Then, standard calculations as in the time series case yield for all $\underline{h}, \underline{k} \in \mathbb{Z}^2$

$$\begin{aligned}
& n^2\, \mathrm{Cov}(\check{\gamma}(\underline{h}), \check{\gamma}(\underline{k})) \\
=\; & \sum_{\underline{r}\in\mathbb{Z}^2 : |r_1|<n, |r_2|<n} \frac{(n-|r_1|)(n-|r_2|)}{n^2} \Bigg(\Big(\gamma(\underline{r}-\underline{k}+\underline{h})\gamma(\underline{r}) + \gamma(\underline{r}+\underline{h})\gamma(\underline{r}-\underline{k}) \Big) \\
& \qquad + (\eta-3) \sum_{\underline{\nu}\in\mathbb{Z}^2} \alpha_{\underline{\nu}}\alpha_{\underline{\nu}-\underline{h}}\alpha_{\underline{\nu}-\underline{r}+\underline{k}-\underline{h}}\alpha_{\underline{\nu}-\underline{r}-\underline{h}}\, \sigma^4 \Bigg) \\
=\; & \sum_{\underline{r}\in\mathbb{Z}^2} \Big(\gamma(\underline{r}-\underline{k}+\underline{h})\gamma(\underline{r}) + \gamma(\underline{r}+\underline{h})\gamma(\underline{r}-\underline{k}) \Big) + (\eta-3)\,\gamma(\underline{h})\gamma(\underline{k}) + o(1) \\
=:\; & V(\underline{h},\underline{k}) + o(1),
\end{aligned}$$

which leads for the sample autocovariances to

$$n^2 \mathrm{Var}\left(\begin{pmatrix} \check{\gamma}(\underline{0}) \\ \check{\gamma}(\underline{h}) \\ \check{\gamma}(\underline{k}) \end{pmatrix} \right) = \begin{pmatrix} V(\underline{0},\underline{0}) & V(\underline{0},\underline{h}) & V(\underline{0},\underline{k}) \\ V(\underline{h},\underline{0}) & V(\underline{h},\underline{h}) & V(\underline{h},\underline{k}) \\ V(\underline{k},\underline{0}) & V(\underline{k},\underline{h}) & V(\underline{k},\underline{k}) \end{pmatrix} + o(1) =: V + o(1).$$

For the quantities $\check{\rho}(\underline{h}) = \check{\gamma}(\underline{h})/\check{\gamma}(\underline{0})$, we get

$$\begin{pmatrix} \check{\rho}(\underline{h}) \\ \check{\rho}(\underline{k}) \end{pmatrix} = \begin{pmatrix} f_1(\check{\gamma}(\underline{0}), \check{\gamma}(\underline{h}), \check{\gamma}(\underline{k})) \\ f_2(\check{\gamma}(\underline{0}), \check{\gamma}(\underline{h}), \check{\gamma}(\underline{k})) \end{pmatrix},$$

where $f(x_1,x_2,x_3) = (f_1(x_1,x_2,x_3), f_2(x_1,x_2,x_3))^T = (x_2/x_1, x_3/x_1)^T$. An application of the Δ-method leads to

$$n^2 \mathrm{Var}\left(\begin{pmatrix} \check{\rho}(\underline{h}) \\ \check{\rho}(\underline{k}) \end{pmatrix} \right) = J_f V J_f^T + o(1),$$

where

$$J_f = \begin{pmatrix} -\gamma(\underline{h})\gamma(\underline{0})^{-2} & \gamma(\underline{0})^{-1} & 0 \\ -\gamma(\underline{k})\gamma(\underline{0})^{-2} & 0 & \gamma(\underline{0})^{-1} \end{pmatrix},$$

such that

$$\begin{aligned}
& n^2 \operatorname{Cov}(\check{\rho}(\underline{h}), \check{\rho}(\underline{k})) \\
= \; & [J_f V J_f^T]_{2,1} + o(1) \\
= \; & [J_f]_{2,\bullet} V [J_f]_{\bullet,1}^T + o(1) \\
= \; & \Big(-\gamma(\underline{k})\gamma(\underline{0})^{-2} V(\underline{0},\underline{0}) + \gamma(\underline{0})^{-1} V(\underline{k},\underline{0})\Big) \Big(-\gamma(\underline{h})\gamma(\underline{0})^{-2}\Big) \\
& + \Big(-\gamma(\underline{k})\gamma(\underline{0})^{-2} V(\underline{0},\underline{h}) + \gamma(\underline{0})^{-1} V(\underline{k},\underline{h})\Big) \gamma(\underline{0})^{-1} + o(1) \\
= \; & \rho(\underline{k})\rho(\underline{h}) \sum_{\underline{r}\in\mathbb{Z}^2} (\rho(\underline{r})\rho(\underline{r}) + \rho(\underline{r})\rho(\underline{r})) - \rho(\underline{h}) \sum_{\underline{r}\in\mathbb{Z}^2} (\rho(\underline{r}+\underline{k})\rho(\underline{r}) + \rho(\underline{r}+\underline{k})\rho(\underline{r})) \\
& -\rho(\underline{k}) \sum_{\underline{r}\in\mathbb{Z}^2} (\rho(\underline{r}-\underline{h})\rho(\underline{r}) + \rho(\underline{r}-\underline{h})\rho(\underline{r})) \\
& + \sum_{\underline{r}\in\mathbb{Z}^2} (\rho(\underline{r}-\underline{h}+\underline{k})\rho(\underline{r}) + \rho(\underline{r}+\underline{k})\rho(\underline{r}-\underline{h})) + o(1) \\
= \; & \sum_{\underline{r}\in\mathbb{Z}^2} \Big\{ 2\rho(\underline{r})^2 \rho(\underline{k})\rho(\underline{h}) - 2\rho(\underline{r}+\underline{k})\rho(\underline{r})\rho(\underline{h}) - 2\rho(\underline{r}-\underline{h})\rho(\underline{r})\rho(\underline{k}) \\
& \qquad + \rho(\underline{r}-\underline{h}+\underline{k})\rho(\underline{r}) + \rho(\underline{r}+\underline{k})\rho(\underline{r}-\underline{h}) \Big\} + o(1).
\end{aligned}$$

In particular, the latter quantity depends exclusively on the second order structure of the linear process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$. □

3 The Vector Autoregressive Sieve Bootstrap

Based on: Meyer, M. and Kreiss, J.-P.:
On the Vector Autoregressive Sieve Bootstrap.
Journal of Time Series Analysis, to appear (2014+).

This chapter deals with the concept of the autoregressive (AR) sieve bootstrap, as introduced in section 1.3, for multivariate stochastic processes. In this context the procedure is called the *vector autoregressive (VAR) sieve bootstrap*; the precise algorithm is stated in the introductory section 3.1. We will investigate its asymptotic behaviour of the procedure following the example of Kreiss, Paparoditis and Politis (2011) for the univariate setting, and derive a similar criterion for checking asymptotic validity.

In section 3.2 we will describe the VAR sieve bootstrap algorithm, and give an overview of the conditions which ensure that an $\mathbb{R}^q$-valued stochastic process $(\underline{X}_t)_{t\in\mathbb{Z}}$ possesses one-sided autoregressive and moving average (MA) representations

$$\underline{X}_t = \sum_{k=1}^{\infty} A_k \, \underline{X}_{t-k} + \underline{\varepsilon}_t, \quad \underline{X}_t = \sum_{k=1}^{\infty} B_k \, \underline{\varepsilon}_{t-k} + \underline{\varepsilon}_t, \tag{3.1}$$

where $(A_k)_{k\in\mathbb{Z}}$ and $(B_k)_{k\in\mathbb{Z}}$ are suitable sequences of $\mathbb{R}^{q\times q}$-matrices, and $(\underline{\varepsilon}_t)_{t\in\mathbb{Z}}$ is the innovation process. Existence of these kinds of representations is crucial for proving bootstrap validity. Here, we will rely heavily on the early work of Wiener and Masani (1957, 1958), who basically showed that a boundedness condition on the determinants of the spectral density matrix of the underlying process $(\underline{X}_t)$ is sufficient to derive representations (3.1).

The first step of the VAR sieve bootstrap algorithm consists of fitting VAR models of finite order p to the given data sample. In order to obtain asymptotic validity, we need the coefficients of these fitted VAR models to converge towards the autoregressive coefficients $(A_k)_{k\in\mathbb{Z}}$ of the underlying process at a certain rate which will be established in section 3.3. This convergence depends largely on a Baxter-inequality for multivariate processes. In contrast to the generalisation to random fields, where deriving this inequality was a major problem as was discussed in chapter 2 of this thesis, the multivariate version of Baxter's inequality is already available in the literature, cf. among others Hannan and Deistler (1988), Theorem 6.6.12. For multivariate processes, the problems with approximating the AR coefficients arise in a different context: Since we are dealing with matrix coefficients instead of real numbers, the commutation property with respect to multiplication is lost, which causes certain arguments from the univariate setting not to be applicable for VAR processes.

After establishing the preliminary results, the main result concerning bootstrap validity will be given in section 3.4. It will show that the VAR sieve bootstrap procedure asymptotically does not mimic the behaviour of the underlying process but the behaviour of a so-called *companion process*, which was introduced by Kreiss, Paparoditis and Politis (2011) for univariate processes. This process has many features in common with the underlying process, including all second order properties. Hence, we can state a general check criterion that can be applied to a large class of statistics: The VAR sieve bootstrap asymptotically works for a specific statistic of interest, if and only if the limiting distributions of the statistic applied to underlying process and applied to the companion process coincide. We will present three examples in section 3.5 that will discuss this criterion.

In section 3.6 we will introduce another possibility of utilizing the VAR sieve bootstrap. For certain non-stationary processes with a deterministic, nonparametric trend function we will show that the procedure is able to approximate the distribution of a trend function estimator. Section 3.7 then contains the proofs of the main theorems, while the proofs of the auxiliary results are deferred to section 3.8.

3.1 The bootstrap procedure and basic notations

Consider a stationary $\mathbb{R}^q$-valued stochastic process $(\underline{X}_t)_{t\in\mathbb{Z}}$ with mean zero and finite second moments. The autocovariance function of $(\underline{X}_t)$ takes its values in $\mathbb{R}^{q\times q}$ and is given by $\Gamma(h) := E(\underline{X}_{t+h}\underline{X}_t^T)$ for lag $h \in \mathbb{Z}$. Throughout this chapter we will denote the entries of q-dimensional time series observations $\underline{X}_t$ by $X_t(1), \dots, X_t(q)$, while the (i,j)-th entry of any matrix A will be denoted by $A^{(i,j)}$.

Let $T_n = T_n(\underline{X}_1, \dots, \underline{X}_n)$ be an estimator for some unknown parameter θ of the process. For an appropriately increasing sequence of real numbers $(c_n)_{n\in\mathbb{N}}$, we assume that the distributions $\mathcal{L}_n = \mathcal{L}(c_n(T_n - \theta))$ converge to a non-degenerated limiting distribution, as $n \to \infty$. Our goal is to approximate the distribution $\mathcal{L}_n$; and we propose using the vector autoregressive sieve bootstrap procedure which works as follows:

The vector autoregressive (VAR) sieve bootstrap algorithm:

(1) Select an order $p = p(n) \in \mathbb{N},\ p \ll n$ and fit a p-th order vector autoregressive model to the given observations, for example by Yule-Walker estimation. Denote by $\widehat{A}_1(p), \dots, \widehat{A}_p(p)$ the estimators of the autoregressive parameters in the fitted model.

(2) Let $\underline{\varepsilon}_t' = \underline{X}_t - \sum_{j=1}^{p} \widehat{A}_j(p)\underline{X}_{t-j},\ t = p+1, \dots, n$, be the residuals of the autoregressive fit and $\widehat{F}_n$ be the empirical distribution function of the centered residuals $\widehat{\underline{\varepsilon}}_t = \underline{\varepsilon}_t' - \overline{\underline{\varepsilon}}$, where $\overline{\underline{\varepsilon}} = (n-p)^{-1} \sum_{t=p+1}^{n} \underline{\varepsilon}_t'$ (in our notation we suppress the dependence of $\underline{\varepsilon}_t'$ and $\widehat{\underline{\varepsilon}}_t$ on p for convenience reasons). Generate a sufficient number of independent random variables $\underline{\varepsilon}_1^*, \underline{\varepsilon}_2^*, \dots$ having identical distribution $\widehat{F}_n$, for example by drawing with replacement from the set of centered residuals. Use these $\underline{\varepsilon}_t^*$ and the parameter estimators to calculate a bootstrap sample $(\underline{X}_1^*, \dots, \underline{X}_n^*)$ according to the generating equation

$$\underline{X}_t^* = \sum_{k=1}^{p} \widehat{A}_k(p)\, \underline{X}_{t-k}^* + \underline{\varepsilon}_t^*.$$

(3) Let $T_{n,(1)}^* = T_n(\underline{X}_1^*, \dots, \underline{X}_n^*)$ be the same estimator as T_n based on the pseudo time series $\underline{X}_1^*, \dots, \underline{X}_n^*$ and θ^* the analogue of θ associated with the bootstrap process $(\underline{X}_t^*)$.

(4) Repeat steps (1)–(3) M times, where M is sufficiently large, in order to obtain independent realisations $T_{n,(1)}^*, \dots, T_{n,(M)}^*$ of the plug-in estimator.

(5) The estimator for $\mathcal{L}_n$ is then given by the empirical distribution of $\mathcal{L}_n^* = \mathcal{L}^*(c_n(T_n^* - \theta^*))$, based on the observations $T_{n,(1)}^*, \dots, T_{n,(M)}^*$.

Here, $\mathcal{L}^*$ and E^* denote probability law and expectation, conditional on the data $\underline{X}_1, \ldots, \underline{X}_n$.

3.2 One-sided representations of multivariate processes

In this section we want to determine conditions which make sure that the process $(\underline{X}_t)$ possesses one-sided autoregressive (AR) and moving average (MA) representations as in (3.1). In order to specify these conditions we first introduce a few notations: We denote by $H_t(\underline{X}) = \overline{sp}\{\underline{X}_s : s \leq t\}$ the closed linear subspace of L^2 that is spanned by the past and present values of the process $\underline{X}$ up to time t. We will restrict ourselves to strictly stationary and purely nondeterministic processes, i.e. $H_{-\infty}(\underline{X}) := \bigcap_{t\in\mathbb{Z}} H_t(\underline{X}) = \{0\}$. Particularly, this implies $\underline{X}_t \notin H_{t-1}(\underline{X})$ which allows us to define the innovation process of $(\underline{X}_t)$: Let $\widehat{\underline{X}}_t$ be the unique projection (in the sense of the Hilbert space L^2) of $\underline{X}_t$ onto the space $H_{t-1}(\underline{X})$. The innovation process $(\underline{\varepsilon}_t)_{t\in\mathbb{Z}}$ of $(\underline{X}_t)$ is then uniquely determined by

$$\underline{\varepsilon}_t = \underline{X}_t - \widehat{\underline{X}}_t \tag{3.2}$$

and it is well known that $(\underline{\varepsilon}_t)$ is a white noise. In accordance with Wiener and Masani (1958) we say that the process $(\underline{X}_t)$ is of *full rank* if the components $\varepsilon_t(1), \ldots, \varepsilon_t(q)$ of $\underline{\varepsilon}_t$ are linearly independent which means that no component can be expressed as a linear combination of the remaining components. Let $\|\cdot\|$ be any matrix norm. Since all matrix norms are equivalent it is not necessary to specify this norm in the following assumptions. However, for convenience reasons, we will use a fixed submultiplicative matrix norm, i.e. $\|AB\| \leq \|A\| \cdot \|B\|$, which fulfils

$$|A^{(i,j)}| \leq \|A\| \leq \sum_{1\leq r,s\leq q} |A^{(r,s)}| \tag{3.3}$$

for each entry $A^{(i,j)}$ for the remainder of this thesis. An example for such a norm is the Frobenius norm. The spectrum of a matrix A, i.e. the set of its eigenvalues, will be denoted by $\sigma(A)$.

We now sum up the assumptions on the underlying process $(\underline{X}_t)$ that we will be working with throughout this chapter:

Assumption 5. *Let $(\underline{X}_t)_{t\in\mathbb{Z}}$ be an $\mathbb{R}^q$-valued, strictly stationary and purely nondeterministic stochastic process of full rank with mean zero and finite second moments (i.e. the second moments of all components are bounded uniformly in t). The autocovariance (matrix) function $\Gamma(\cdot)$ of $(\underline{X}_t)$ fulfils $\sum_{h=-\infty}^{\infty}(1+|h|)^r \|\Gamma(h)\| < \infty$ for some $r \geq 0$ that will be specified in the respective results later on. The spectral density matrix $W(\cdot)$ of $(\underline{X}_t)$ fulfils the so-called boundedness condition, cf. Wiener and Masani (1958): There exists a constant $c > 0$ such that*

$$\min(\sigma(W(\lambda))) \geq c$$

for all frequencies $\lambda \in (-\pi, \pi]$, i.e. the eigenvalues of the spectral density matrix are uniformly bounded away from zero.

Note that the summability condition on the autocovariances ensures that the eigenvalues of $W(\lambda)$ are also bounded from above, uniformly for all frequencies $\lambda \in (-\pi, \pi]$.

We will now derive a certain autoregressive representation for all processes $(\underline{X}_t)_{t\in\mathbb{Z}}$ that fulfil Assumption 5. First, following Wiener and Masani (1957), Lemma 6.9, the innovation process $(\underline{\varepsilon}_t)_{t\in\mathbb{Z}}$ of $(\underline{X}_t)$ is an uncorrelated white noise process with covariance matrix $\mathrm{Cov}(\underline{\varepsilon}_s, \underline{\varepsilon}_t) = \delta_{s,t}\Sigma$. Because of the full-rank assumption, Σ is an invertible Gramian matrix, cf. Wiener and Masani (1957), page 136, and therefore positive definite. Since both Σ and Σ^{-1} are symmetric, the positive-definiteness of Σ implies that all eigenvalues of both Σ and Σ^{-1} are strictly positive. This allows us to take the principal square root $\Sigma^{1/2}$ of Σ in the sense that $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$, as well as the square root $\Sigma^{-1/2}$ of Σ^{-1}. By $\underline{\varepsilon}_t^{\mathrm{nor}} := \Sigma^{-1/2}\underline{\varepsilon}_t$ we can define the *normalized innovation process* $(\underline{\varepsilon}_t^{\mathrm{nor}})_{t\in\mathbb{Z}}$, which is obviously an uncorrelated white noise with $\mathrm{Cov}(\underline{\varepsilon}_s^{\mathrm{nor}}, \underline{\varepsilon}_t^{\mathrm{nor}}) = \delta_{s,t}I$, where I denotes the identity matrix.

According to Wiener and Masani (1958), Theorems 5.5 and 5.7, under Assumption 5, there exist sequences of coefficient matrices $(C_k)_{k\in\mathbb{N}}$ and $(D_k)_{k\in\mathbb{N}}$, not depending on $t \in \mathbb{Z}$, such that the projection of $\underline{X}_t$ on its infinite past and the normalized innovation process possess the representations

$$\widehat{\underline{X}}_t = \sum_{k=1}^{\infty} C_k\, \underline{\varepsilon}_{t-k}^{\mathrm{nor}}, \quad \text{and} \quad \underline{\varepsilon}_t^{\mathrm{nor}} = \sum_{k=0}^{\infty} D_k\, \underline{X}_{t-k} \tag{3.4}$$

for all $t \in \mathbb{Z}$, where $D_0 = \Sigma^{-1/2}$. The infinite sums in (3.4) are L_2-convergent. Therefore, by setting $A_0 := \Sigma^{1/2} D_0 = I$, $A_k := -\Sigma^{1/2} D_k$ and $B_k := C_k\, \Sigma^{-1/2}$ for all

$k \in \mathbb{N}$, we can derive from (3.4) and (3.2) the following autoregressive and moving average representations of the process $(\underline{X}_t)$:

$$\underline{X}_t = \sum_{k=1}^{\infty} A_k \, \underline{X}_{t-k} + \underline{\varepsilon}_t, \quad \text{and} \quad \underline{X}_t = \sum_{k=1}^{\infty} B_k \, \underline{\varepsilon}_{t-k} + \underline{\varepsilon}_t \tag{3.5}$$

We now want to deduce some summability properties for the coefficient matrices $(A_k)_{k\in\mathbb{N}}$ and $(B_k)_{k\in\mathbb{N}}$. For this reason, we introduce the weight function $\nu(\cdot)$ and define $\nu(j) := (1+|j|)^r$ for all $j \in \mathbb{Z}$ and a fixed $r \geq 0$. Clearly, from $|j| \leq |j-k|+|k|$ for all $j, k \in \mathbb{Z}$ we get

$$\nu(j) = (1+|j|)^r \leq (1+|j-k|+|k|+|k|\cdot|j-k|)^r = \nu(k)\,\nu(j-k). \tag{3.6}$$

For all weight functions of this type, Cheng and Pourahmadi (1993) define a class of matrix functions in the following way: Let $\mathcal{C}_\nu$ be the class of all $q \times q$-matrix-valued functions on $(-\pi, \pi]$ with $\mathbb{C}^{q\times q}$-valued Fourier coefficient matrices $(F_k)_{k\in\mathbb{Z}}$, fulfilling

$$\sum_{h=-\infty}^{\infty} \nu(|h|)\,\|F_h\| < \infty.$$

Using the particular weight function from above, we see that under Assumption 5 the spectral density matrix function $W(\cdot)$ is in $\mathcal{C}_\nu$ because its Fourier coefficients are the autocovariance matrices $\Gamma(h)$. Furthermore, under the imposed conditions, W can be decomposed as

$$W(\lambda) = \phi(\lambda)\,\phi(\lambda)^*$$

where * denotes the conjugate transpose and ϕ is called the optimal factor of W, cf. Cheng and Pourahmadi (1993), p. 116, and also Wiener and Masani (1958), p. 121. The boundedness condition ensures that $\det W(\lambda) > 0$ for all λ and it holds

$$\begin{aligned} |\det \phi(\lambda)|^2 &= \det \phi(\lambda)\,\overline{(\det \phi(\lambda)^T)} \\ &= \det(\phi(\lambda)\phi(\lambda)^*) = \det W(\lambda) \neq 0 \end{aligned} \tag{3.7}$$

and therefore, $\phi(\lambda)$ is invertible for all λ. From $W \in \mathcal{C}_\nu$ and Theorem 1.1 in Cheng and Pourahmadi (1993) it follows that $\phi \in \mathcal{C}_\nu$. With (3.7) we get from Cheng and Pourahmadi (1993), page 117, that $\phi^{-1} \in \mathcal{C}_\nu$, as $\mathcal{C}_\nu$ is a Banach algebra. The sequences of matrices (C_k) and (D_k) from (3.4), together with some C_0 that does not show up in representation (3.4), are the Fourier coefficients of the functions ϕ and ϕ^{-1}, respectively, i.e.

$$\phi(\lambda) = \sum_{k=0}^{\infty} C_k\, e^{ik\lambda} \quad \text{and} \quad \phi^{-1}(\lambda) = \sum_{k=0}^{\infty} D_k\, e^{ik\lambda} \tag{3.8}$$

cf. Wiener and Masani (1958), Theorems 5.5, 5.7, equation (5.6) and Definition 2.6. Combining these results yields

$$\sum_{k=0}^{\infty}(1+|k|)^r \, \|C_k\| < \infty \quad \text{and} \quad \sum_{k=0}^{\infty}(1+|k|)^r \, \|D_k\| < \infty.$$

With the definition of the matrices A_k and the submultiplicative property of the matrix norm it is easy to derive

$$\begin{aligned}\sum_{k=1}^{\infty}(1+|k|)^r \, \|A_k\| &= \sum_{k=1}^{\infty}(1+|k|)^r \, \|\Sigma^{1/2}\Sigma^{-1/2}\,(-A_k)\| \\ &\leq \|\Sigma^{1/2}\| \sum_{k=1}^{\infty}(1+|k|)^r \, \|D_k\| < \infty.\end{aligned}$$

Analogously, the same holds true for the coefficients B_k in connection with C_k and we can conclude that the moving average and autoregressive coefficients from (3.5) have the following summability properties:

$$\sum_{k=1}^{\infty}(1+|k|)^r \, \|A_k\| < \infty \quad \text{and} \quad \sum_{k=1}^{\infty}(1+|k|)^r \, \|B_k\| < \infty. \tag{3.9}$$

We finish this section with another preliminary result. From the assertions above we get that the functions ϕ and ϕ^{-1} are equal to their (absolutely convergent) Fourier series as in (3.8). As for all absolutely convergent series, the product of the two series is given by

$$\begin{aligned}I = \phi(\lambda)\,\phi^{-1}(\lambda) &= \sum_{k=0}^{\infty} C_k\, e^{ik\lambda} \cdot \sum_{k=0}^{\infty} D_k\, e^{ik\lambda} \\ &= \sum_{k=0}^{\infty}\left(\sum_{j=0}^{k} C_k D_{k-j}\right) e^{ik\lambda}.\end{aligned} \tag{3.10}$$

Because equality in (3.10) holds for all $\lambda \in (-\pi, \pi]$ it follows immediately that the coefficients of the Cauchy product on the right-hand side are trivial, i.e.

$$\sum_{j=0}^{k} C_k D_{k-j} = \begin{cases} I & , \text{ for } k = 0 \\ 0 & , \text{ for all } k > 0 \end{cases} \tag{3.11}$$

where 0 denotes the $q \times q$-zero-matrix. The functions ϕ and ϕ^{-1} are defined on $(-\pi, \pi]$, but of course we can interpret them as functions on the complex unit circle. To avoid confusion in the notation we will denote these functions by

$$\widetilde{\phi}(e^{i\lambda}) := \phi(\lambda) \quad \text{and} \quad \widetilde{\phi}^{-1}(e^{i\lambda}) := \phi^{-1}(\lambda)$$

which defines the functions $\widetilde{\phi}$ and $\widetilde{\phi}^{-1}$ for all $|z| = 1$. As the Fourier series of ϕ and ϕ^{-1} are absolutely convergent we can extend the domain of $\widetilde{\phi}$ and $\widetilde{\phi}^{-1}$ to the entire closed unit disk via

$$\widetilde{\phi}(z) := \sum_{k=0}^{\infty} C_k \, z^k \quad \text{and} \quad \widetilde{\phi}^{-1}(z) := \sum_{k=0}^{\infty} D_k \, z^k$$

where it is easy to see that the two power series are absolutely convergent for all $|z| \leq 1$. Using (3.11) we immediately get $\widetilde{\phi}(z)\,\widetilde{\phi}^{-1}(z) = I$ for all $|z| \leq 1$ which justifies the notation $\widetilde{\phi}^{-1}$. Furthermore, this implies

$$\det \widetilde{\phi}(z) \neq 0 \quad \text{and} \quad \det \widetilde{\phi}^{-1}(z) \neq 0 \qquad \text{for all } |z| \leq 1. \tag{3.12}$$

Corresponding to (3.5) we define the autoregressive polynomial of the underlying process as

$$A(z) := I - \sum_{k=1}^{\infty} A_k z^k,$$

which is (in the sense of our matrix norm $\|\cdot\|$) an absolutely convergent power series for all $|z| \leq 1$ according to (3.9). Since $\widetilde{\phi}^{-1}(z) = \Sigma^{-1/2} A(z)$ for all $|z| \leq 1$ and $\Sigma^{-1/2}$ is regular, we immediately get from (3.12) that

$$\det\, A(z) \neq 0 \quad \forall\, |z| \leq 1. \tag{3.13}$$

We conclude this section by summarizing that all processes fulfilling Assumption 5 possess the one-sided representations

$$\underline{X}_t = \sum_{k=1}^{\infty} A_k \, \underline{X}_{t-k} + \underline{\varepsilon}_t, \quad \text{and} \quad \underline{X}_t = \sum_{k=1}^{\infty} B_k \, \underline{\varepsilon}_{t-k} + \underline{\varepsilon}_t,$$

where the coefficient matrices A_k, B_k fulfil the summability conditions (3.9).

3.3 Convergence of finite predictor coefficients

The one-sided autoregressive representations from the previous section are based on the projection of $\underline{X}_t$ on its infinite past. We will now turn our attention to finite predictors, i.e. projections on the finite past. It is well known that for $p \in \mathbb{N}$ the L^2-projection of $\underline{X}_t$ on $\overline{sp}\{\underline{X}_{t-1}, \ldots, \underline{X}_{t-p}\}$ can be derived by finding $q \times q$-matrices $A_1(p), \ldots, A_p(p)$ such that

$$\underline{X}_t - \sum_{k=1}^{p} A_k(p)\, \underline{X}_{t-k}$$

is orthogonal to $\underline{X}_{t-j}$ for $j = 1, \dots, p$ in the L^2-sense. Solving this problem leads to the well-known Yule-Walker equations for multivariate processes, cf. Brockwell and Davis (1991), p. 411:

$$[\, A_1(p), \dots, A_p(p)\,] \;\; G(p) = [\, \Gamma(1), \dots, \Gamma(p)\,] \tag{3.14}$$

where $G(p) \in \mathbb{R}^{pq \times pq}$ is given by

$$G(p) := \begin{bmatrix} \Gamma(0) & \cdots & \Gamma(p-1) \\ \vdots & \ddots & \vdots \\ \Gamma(-p+1) & \cdots & \Gamma(0) \end{bmatrix}.$$

Since we assume in Assumption 5 that our process has full rank, we get from Wiener and Masani (1958), p. 101, that the matrix $G(p)$ is invertible and, thus, the system (3.14) has a unique solution. We call the solution $A_1(p), \dots, A_p(p)$ the *finite predictor coefficients* of the process and get

$$[\, A_1(p), \dots, A_p(p)\,] = [\, \Gamma(1), \dots, \Gamma(p)\,] \;\; G(p)^{-1}.$$

It is of critical importance for our sieve bootstrap scheme that these finite predictor coefficients converge towards the sequence of autoregressive coefficients $(A_k)_{k \in \mathbb{N}}$ of the underlying process, as given by (3.5), as p tends to infinity. We get this convergence from a multivariate version of Baxter's inequality which is basically taken from Hannan and Deistler (1988), Theorem 6.6.12. Our proof merely ensures that the required conditions of the proof of Hannan and Deistler are fulfilled under the imposed conditions from Assumption 5. The result then reads:

Lemma 3.1. (Baxter's Inequality) *Let $(\underline{X}_t)_{t \in \mathbb{Z}}$ be a process that fulfils Assumption 5 for some $r \geq 0$. Let $A_1(p), \dots, A_p(p)$ be its finite predictor coefficients as defined above and $(A_k)_{k \in \mathbb{N}}$ its autoregressive coefficients as in (3.5). Then there exist constants $p_0 \in \mathbb{N}$ and $C < \infty$ such that*

$$\sum_{k=1}^{p} (1 + |k|)^r \, \|A_k(p) - A_k\| \leq C \cdot \sum_{k=p+1}^{\infty} (1 + |k|)^r \, \|A_k\| \qquad \forall\, p \geq p_0 \tag{3.15}$$

and the right-hand side converges to zero as $p \to \infty$.

The convergence of the autoregressive coefficients in Baxter's inequality, cf. (3.15), is closely related to a similar convergence of moving average parameters. In order to state such a result, we first make sure that autoregressive models determined by the finite predictor coefficients $A_1(p), \dots, A_p(p)$ can be expressed as moving average processes of possibly infinite order. For each stationary q-variate white noise process $(\underline{e}_t)_{t\in\mathbb{Z}}$ with finite variances we can define an AR(p)-process $(\underline{Y}_t)_{t\in\mathbb{Z}}$, say, by

$$\underline{Y}_t := \sum_{k=1}^{p} A_k(p)\, \underline{Y}_{t-k} + \underline{e}_t,$$

where $A_1(p), \dots, A_p(p)$ are the finite predictor coefficients. If the autoregressive polynomial $A_p(z) := I - \sum_{k=1}^{p} A_k(p)\, z^k$ fulfils

$$\det A_p(z) \neq 0 \qquad \forall\, |z| \leq 1,$$

it is well-known that $(\underline{Y}_t)$ is stable and can be expressed as a one-sided moving average process of infinte order with the same white noise process, i.e.

$$\underline{Y}_t = \sum_{k=1}^{\infty} B_k(p)\, \underline{e}_{t-k} + \underline{e}_t.$$

The moving average coefficients $(B_k(p))_{k\in\mathbb{N}}$ are the ones that appear in the power series expansion

$$A_p(z)^{-1} = I + \sum_{k=1}^{\infty} B_k(p)\, z^k \qquad \forall\, |z| \leq 1, \tag{3.16}$$

cf. Brockwell and Davis (1991), Theorem 11.3.1 and its proof. Our next preliminary result will not only guarantee the invertibility of $A_p(z)$, but also ensure that $\det A_p(z)$ is uniformly bounded away from zero on the entire closed unit disk plus on a small ring around the disk. This assertion will be needed to prove our remaining auxiliary results.

Lemma 3.2. *Let $(\underline{X}_t)_{t\in\mathbb{Z}}$ be a process that fulfils Assumption 5 and $A_1(p), \dots, A_p(p)$ be its finite predictor coefficients, i.e. the solution of* (3.14)*. Let $A_p(z) := I - \sum_{k=1}^{p} A_k(p)\, z^k$. Then there exist $p_1 \in \mathbb{N}$ and $\delta > 0$ such that*

$$\inf_{|z|\leq 1+(1/p)} |\det\, A_p(z)| \geq \delta \qquad \forall\, p \geq p_1.$$

It is worth noting that the proof shows $\det\, A_p(z) \neq 0$ for all $|z| \leq 1$ and for all $p \in \mathbb{N}$, whereas the stronger condition of the previous Lemma only holds for $p \geq p_1$. Therefore, the power series expansion (3.16) is valid for all $p \in \mathbb{N}$.

We can now state the aforementioned result about the convergence of the moving average parameters:

Lemma 3.3. *Let $(\underline{X}_t)_{t\in\mathbb{Z}}$ be a process that fulfils Assumption 5. Let $(A_k)_{k\in\mathbb{N}}$ and $(B_k)_{k\in\mathbb{N}}$ be its autoregressive and moving average coefficients as in (3.5). Let $p_1 \in \mathbb{N}$ be the constant defined in Lemma 3.2 and $(B_k(p))_{k\in\mathbb{N}}$ be the power series coefficients of $A_p(z)^{-1}$ as defined in (3.16). Then there exists $p_2 \geq p_1$ and a constant $C < \infty$ such that*

$$\sum_{k=1}^{\infty}(1+|k|)^r \left\|B_k(p) - B_k\right\| \leq \; C \cdot \sum_{k=p+1}^{\infty} (1+|k|)^r \left\|A_k\right\| \qquad \forall\, p \geq p_2.$$

Of course, if we want to apply the AR sieve bootstrap procedure to a given data sample $(\underline{X}_1, \ldots, \underline{X}_n)$, we cannot use the coefficients $A_1(p), \ldots, A_p(p)$ since they are unknown. Instead, we replace them by the parameter estimators $\widehat{A}_1(p), \ldots, \widehat{A}_p(p)$. One common choice to obtain these estimators is the Yule-Walker method, which uses the simple idea to replace the autocovariances in (3.14) by the empirical autocovariances $\widehat{\Gamma}(h)$, which are given by

$$\widehat{\Gamma}(h) = \begin{cases} \frac{1}{n}\sum_{t=1}^{n-h}(\underline{X}_{t+h} - \overline{\underline{X}})(\underline{X}_t - \overline{\underline{X}})^T, & \text{for } 0 \leq h \leq n-1, \\ \frac{1}{n}\sum_{t=-h+1}^{n}(\underline{X}_{t+h} - \overline{\underline{X}})(\underline{X}_t - \overline{\underline{X}})^T, & \text{for } -n+1 \leq h < 0, \end{cases} \tag{3.17}$$

where $\overline{\underline{X}} = n^{-1}\sum_{t=1}^{n} \underline{X}_t$ denotes the sample mean. The Yule-Walker estimators can then be obtained from solving the linear system

$$\left[\widehat{A}_1(p), \ldots, \widehat{A}_p(p)\right] \; \widehat{G}(p) = \left[\widehat{\Gamma}(1), \ldots, \widehat{\Gamma}(p)\right], \tag{3.18}$$

where

$$\widehat{G}(p) := \begin{bmatrix} \widehat{\Gamma}(0) & \cdots & \widehat{\Gamma}(p-1) \\ \vdots & \ddots & \vdots \\ \widehat{\Gamma}(-p+1) & \cdots & \widehat{\Gamma}(0) \end{bmatrix}.$$

This raises the question whether $\widehat{G}(p)$ is invertible. The following Lemma gives a characterization of the invertibility property.

Lemma 3.4. *Let $1 \leq p < n$ and $(\underline{X}_1, \ldots, \underline{X}_n)$ be any data sample. Denote the sample mean by $\overline{\underline{X}} := n^{-1}\sum_{t=1}^{n}\underline{X}_t$ and define the centered observations $\underline{Y}_j := \underline{X}_j - \overline{\underline{X}}$, $j \in \{1, \ldots, n\}$, and the $qp \times (n+p-1)$-matrix*

$$H = \begin{pmatrix} & & & \underline{Y}_n & \cdots & \cdots & \cdots & \cdots & \underline{Y}_1 \\ & & \underline{Y}_n & \cdots & \cdots & \cdots & \cdots & \underline{Y}_1 & \\ & \iddots & \iddots & \iddots & \iddots & \iddots & \iddots & & \\ \underline{Y}_n & \cdots & \cdots & \cdots & \cdots & \underline{Y}_1 & & & \end{pmatrix}.$$

Then the following two assertions are equivalent:

(i) The matrix $\widehat{G}(p)$ is invertible.

(ii) It holds $qp \leq n+p-1$, and the rows of the matrix H are linearly independent.

Remark 3.5. Consider the univariate setting $q = 1$ as a special case. Lemma 3.4 then states that $\widehat{G}(p)$ is invertible if and only if the following two conditions are fulfilled: Firstly, we need $p < n$ and $1 \leq p \leq n + p - 1$, i.e. $n > 1$, which is trivial. Secondly, the rows of the matrix

$$H = \begin{pmatrix} & & & & & Y_n & \cdots & \cdots & \cdots & \cdots & Y_1 \\ & & & & Y_n & \cdots & \cdots & \cdots & \cdots & Y_1 & \\ & \iddots & \iddots & \iddots & \iddots & \iddots & \iddots & & & & \\ Y_n & \cdots & \cdots & \cdots & \cdots & Y_1 & & & & & \end{pmatrix}$$

have to be linearly independent. Taking an arbitrary linear combination of the rows of the matrix and setting it equal to zero immediately yields: the trivial linear combination is the only solution unless it holds $Y_1 = \ldots = Y_n = 0$ which is equivalent to $X_1 = \ldots = X_n$. This is a well-known characterization for uniqueness of the Yule-Walker estimators for univariate time series. □

Remark 3.6. In the multivariate case, linear dependence of the rows of the matrix H in Lemma 3.4 means that there exists a vector $\underline{a} = (\underline{a}_1^T, \ldots, \underline{a}_p^T)^T$, $\underline{a}_i \in \mathbb{R}^q$ for all i, such that $\underline{a}^T H = \underline{0}^T$. Looking particularly at the interior columns of H, it follows:

$$\underline{a}_1^T \, \underline{Y}_{n-k} + \ldots + \underline{a}_p^T \, \underline{Y}_{n-k-p+1} = 0 \qquad \forall \, k \in \{0, \ldots, n-p\}$$

Setting $c := \sum_{j=1}^{p} \underline{a}_j^T \, \overline{\underline{X}}_n$, this is equivalent to

$$\underline{a}_1^T \, \underline{X}_{n-k} + \ldots + \underline{a}_p^T \, \underline{X}_{n-k-p+1} = c \qquad \forall \, k \in \{0, \ldots, n-p\},$$

i.e. there is an exact linear relation within each string of p consecutive observations in our sample. To be precise, the very same linear relation, defined by the constant vector $\underline{a}$, links the first p observations, as well as the last p ones and all other subsamples of length p in between. It is obvious that such a situation is a degenerate case which will not occur for an actual data sample for $p \ll n$. Thus, it is reasonable to assume that $\widehat{G}(p)$ is invertible. □

We will now turn our attention to the convergence of the parameter estimators $\widehat{A}_1(p), \ldots, \widehat{A}_p(p)$ towards the finite predictor coefficients $A_1(p), \ldots, A_p(p)$. At first we will impose an assumption on this convergence rate and then give an example for conditions that are sufficient for this assumption to be satisfied. Note that the order p of the autoregressive fit in our bootstrap scheme actually depends on the sample size n and has to be chosen properly. In the following we will write $p = p(n)$ and only (at times) suppress the dependence on n in order to simplify the notation.

Assumption 6. *It holds $p = p(n) \to \infty$, as $n \to \infty$. Furthermore, $p(n)$ increases slowly enough such that*

$$p(n)^2 \cdot \sum_{k=1}^{p(n)} \|\widehat{A}_k(p(n)) - A_k(p(n))\| = \mathcal{O}_P(1).$$

Remark 3.7. Note that

$$p(n)^2 \cdot \sum_{k=1}^{p(n)} \|\widehat{A}_k(p(n)) - A_k(p(n))\| \leq p(n)^3 \cdot \sup_{1\leq k\leq p(n)} \|\widehat{A}_k(p(n)) - A_k(p(n))\|. \quad (3.19)$$

Under the imposed conditions the right-hand side can be bounded with Theorem 2.1 in Hannan and Kavalieris (1986) via

$$\sup_{1\leq k\leq p(n)} \|\widehat{A}_k(p(n)) - A_k(p(n))\| = \mathcal{O}_P((\ln n/n)^{1/2}) + o(1) \sum_{k=p(n)+1}^{\infty} \|A_k\|. \quad (3.20)$$

Furthermore, assume that it holds Assumption 5 with $r = 3$. From (3.9) it is then easy to see that

$$\begin{aligned} p(n)^3 \sum_{k=p(n)+1}^{\infty} \|A_k\| &= \sum_{k=p(n)+1}^{\infty} p(n)^3 \|A_k\| \\ &\leq \sum_{k=p(n)+1}^{\infty} k^3 \|A_k\| \leq \sum_{k=1}^{\infty} k^3 \|A_k\| = \mathcal{O}(1). \end{aligned}$$

Combining this bound with (3.19) and (3.20), it follows that the assertion from Assumption 6 holds if $p(n)^3 (\ln n/n)^{1/2} = \mathcal{O}(1)$. Hence, we get the desired rate of convergence if we choose $p = p(n)$ increasing as $n \to \infty$, but slowly enough, namely $p(n) = \mathcal{O}\Big((n/\ln n)^{1/6}\Big)$. □

The convergence of the Yule-Walker estimators in Assumption 6 also carries over to the corresponding moving average representations. Since $\widehat{A}_p(z) := I - \sum_{k=1}^{p} \widehat{A}_k(p)\, z^k$ is invertible for all $|z| \leq 1$, cf. Brockwell and Davis (1991), p. 419, its inverse possesses the power series expansion

$$\widehat{A}_p(z)^{-1} = I + \sum_{k=1}^{\infty} \widehat{B}_k(p)\, z^k \qquad \forall\, |z| \leq 1. \quad (3.21)$$

Therefore, considering the construction of the bootstrap process $(\underline{X}_t^*)$ in step (2) of the autoregressive sieve bootstrap procedure, $(\underline{X}_t^*)$ possesses the moving average representation

$$\underline{X}_t^* = \sum_{k=1}^{\infty} \widehat{B}_k(p)\, \underline{\varepsilon}_{t-k}^* + \underline{\varepsilon}_t^*. \tag{3.22}$$

Our final preliminary result makes sure that the difference between the coefficients $(\widehat{B}_k(p))_{k\in\mathbb{N}}$ and $(B_k(p))_{k\in\mathbb{N}}$ becomes asymptotically small in probability for each $k \in \mathbb{N}$:

Lemma 3.8. *Let $(\underline{X}_t)_{t\in\mathbb{Z}}$ be a process that fulfils Assumptions 5 and 6. Let $(B_k(p))_{k\in\mathbb{N}}$ be the power series coefficients of $A_p(z)^{-1}$ as defined in (3.16) and $(\widehat{B}_k(p))_{k\in\mathbb{N}}$ be the power series coefficients of $\widehat{A}_p(z)^{-1}$ as in (3.21). Then there exists $p_3 \in \mathbb{N}$ such that it holds uniformly in $k \in \mathbb{N}$ and for all $p \geq p_3$:*

$$\|\widehat{B}_k(p) - B_k(p)\| \leq \left(1 + \frac{1}{p}\right)^{-k} \frac{1}{p^2}\, \mathcal{O}_P(1).$$

The proofs of all lemmas from this section can be found in section 3.8.

3.4 Asymptotic results

In this section we will show that the VAR sieve bootstrap procedure, applied to a class of statistics that will be specified later on, asymptotically does not mimic the behaviour of the underlying process but the one of a slightly different process which we will call the companion process of $(\underline{X}_t)$. The concept of the companion process was introduced by Kreiss, Paparoditis and Politis (2011) for univariate processes and translates naturally to the multivariate setting. We define an i.i.d. white noise process $(\widetilde{\underline{\varepsilon}}_t)_{t\in\mathbb{Z}}$ where $\mathcal{L}(\widetilde{\underline{\varepsilon}}_t) = \mathcal{L}(\underline{\varepsilon}_t)$ for any $t \in \mathbb{Z}$ (note that $(\underline{\varepsilon}_t)_{t\in\mathbb{Z}}$ is strictly stationary, i.e. $(\widetilde{\underline{\varepsilon}}_t)_{t\in\mathbb{Z}}$ is well-defined). The *companion process* of $(\underline{X}_t)$ is then given by

$$\widetilde{\underline{X}}_t = \sum_{k=1}^{\infty} A_k\, \widetilde{\underline{X}}_{t-k} + \widetilde{\underline{\varepsilon}}_t, \quad t \in \mathbb{Z}, \tag{3.23}$$

where the coefficient matrices $(A_k)_{k\in\mathbb{N}}$ are exactly the ones from the autoregressive representation (3.5) of the underlying process $(\underline{X}_t)$. Therefore, the companion process also possesses the moving average representation

$$\widetilde{\underline{X}}_t = \sum_{k=1}^{\infty} B_k\, \widetilde{\underline{\varepsilon}}_{t-k} + \widetilde{\underline{\varepsilon}}_t, \quad t \in \mathbb{Z}, \tag{3.24}$$

with coefficients $(B_k)_{k\in\mathbb{N}}$ as in (3.5). The only difference between $(\underline{X}_t)$ and $(\underline{\widetilde{X}}_t)$ is the dependence structure of the respective noise processes $(\underline{\varepsilon}_t)$ and $(\underline{\widetilde{\varepsilon}}_t)$. While $(\underline{\widetilde{\varepsilon}}_t)$ is i.i.d., $(\underline{\varepsilon}_t)$ is strictly stationary but the random vectors $\underline{\varepsilon}_s$ and $\underline{\varepsilon}_t$ are only uncorrelated for $s \neq t$. Nevertheless, it is easy to see from (3.24) that all second order properties of $(\underline{X}_t)$ and $(\underline{\widetilde{X}}_t)$ are identical, i.e. the two processes possess identical autocovariances and spectral densities.

In the following assumption we will specify the class of statistics for which the check criterion from the upcoming theorem will be applicable. This class is commonly referred to as *functions of generalized means*, and has been introduced by Künsch (1989).

Assumption 7. *Let*

$$T_n = f\left(\frac{1}{n-m+1}\sum_{t=1}^{n-m+1} g(\underline{X}_t,\ldots,\underline{X}_{t+m-1})\right)$$

for some $m \in \{1,\ldots,n\}$ and functions $g : \mathbb{R}^{mq} \to \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$, where $d \geq 1$, fulfilling the following smoothness assumptions: f is continuously differentiable in a neighborhood of $\underline{\theta} := E\Big(g(\underline{X}_t,\ldots,\underline{X}_{t+m-1})\Big)$, and the gradient of f at $\underline{\theta}$ does not vanish, i.e.

$$\nabla f(\underline{\theta}) = \left(\frac{\partial f(\underline{x})}{\partial x_1},\ldots,\frac{\partial f(\underline{x})}{\partial x_d}\right)\Bigg|_{\underline{x}=\underline{\theta}} \neq (0,\ldots,0).$$

For some $h \geq 1$ all component functions $g_1,\ldots,g_d$ of g are h times continuously differentiable and all h-th-order derivatives satisfy a Lipschitz condition, i.e. for all $i = 1,\ldots,d$ and for all $(h_1,\ldots,h_{mq}) \in \mathbb{N}_0^{mq}$ with $\sum_{u=1}^{mq} h_u = h$ the derivative

$$\frac{\partial^h g_i(\underline{x})}{\partial^{h_1}x_1\ldots\partial^{h_{mq}}x_{mq}}$$

is Lipschitz.

Remark 3.9. We will explain the conditions of the previous assumption in more detail: The class of statistics from Assumption 7 contains, among other things, the sample mean and versions of the sample autocovariance and sample autocorrelation. To obtain the latter two statistics, one typically uses a function g which is not Lipschitz. For example, in the case of the sample autocovariance between the first and q-th component of $\underline{X}_t$ at lag h, i.e. $\widehat{\Gamma}(h)^{(q,1)}$, one may choose $m = h+1$ and $g(x_1,\ldots,x_{mq}) = x_1 x_{mq}$. Then T_n from Assumption 7 translates to taking the empirical mean of the observations $\underline{X}_{t+h}(q) \cdot \underline{X}_t(1)$. Now observe that g itself is *not Lipschitz*, but all of its first order partial derivatives are. This is the why we allow for non-Lipschitz functions g in Assumption 7, and merely assume that there exists a number $1 \leq h < \infty$ such that all derivatives of order h (but *not* up to order h) are Lipschitz. $\square$

In order to state the validity theorem, we define for any T_n fulfilling Assumption 7 $\widetilde{T}_n$ as the statistic T_n applied to a sample from the companion process, i.e.

$$\widetilde{T}_n = f\left(\frac{1}{n-m+1}\sum_{t=1}^{n-m+1} g(\widetilde{\underline{X}}_t, \ldots, \widetilde{\underline{X}}_{t+m-1})\right).$$

The result holds under the following assumptions on empirical quantities of the underlying process:

Assumption 8. *For all continuity points* $\underline{x} \in \mathbb{R}^q$ *of the distribution function* F *of* $\mathcal{L}(\underline{\varepsilon}_1)$ *it holds*

$$F_n(\underline{x}) \xrightarrow{P} F(\underline{x}),$$

where F_n *denotes the empirical distribution function of* $\underline{\varepsilon}_1, \ldots, \underline{\varepsilon}_n$. *Furthermore, for each* $j \in \{1, \ldots, q\}$, *we have* $E\Big(\varepsilon_1(j)^{2(h+2)}\Big) < \infty$ *as well as the following convergence of empirical moments towards their theoretical counterparts:*

$$\frac{1}{n}\sum_{t=1}^{n} \varepsilon_t(j)^{2w} \xrightarrow{P} E\Big(\varepsilon_1(j)^{2w}\Big) \qquad \forall\, w \leq h+2,$$

where $\varepsilon_t(1), \ldots, \varepsilon_t(q)$ *are the components of* $\underline{\varepsilon}_t$ *and* h *is the constant specified in Assumption 7.*

Theorem 3.10. *Let* $(\underline{X}_t)_{t\in\mathbb{Z}}$ *be a process fulfilling Assumption 5 for* $r = 1$. *Assume Assumptions 6, 7 and 8 hold. Then, for* $\widetilde{T}_n$ *as defined above and* $T_n^* = T_n(\underline{X}_1^*, \ldots, \underline{X}_n^*)$, *it holds*

$$d_K\left(\mathcal{L}^*\left(\sqrt{n}(T_n^* - f(\underline{\theta}^*))\right), \mathcal{L}\left(\sqrt{n}(\widetilde{T}_n - f(\widetilde{\underline{\theta}}))\right)\right) = o_P(1)$$

as $n \to \infty$, *where* $\underline{\theta}^* = E^*(g(\underline{X}_t^*, \ldots, \underline{X}_{t+m-1}^*))$, $\widetilde{\underline{\theta}} = E(g(\widetilde{\underline{X}}_t, \ldots, \widetilde{\underline{X}}_{t+m-1}))$ *and* d_K *denotes the Kolmogorov distance.*

The proof can be found in section 3.7. This result shows for all statistics from Assumption 7 that the VAR sieve bootstrap procedure asymptotically approximates the distribution $\widetilde{T}_n$ instead of the one of T_n. Therefore, the bootstrap procedure works asymptotically *if and only if* the limiting distributions of T_n and $\widetilde{T}_n$ coincide. We will give a few examples of the application of this check criterion in the following section.

The proof of Theorem 3.10 will be based on the following auxiliary results:

Lemma 3.11. *Under the assumptions of Theorem 3.10, the following assertions hold true:*

- $$\sum_{j=1}^{\infty} \| \widehat{B}_j(p) - B_j \| \xrightarrow{P} 0, \tag{3.25}$$
- $$\sum_{j=1}^{\infty} j\, \|\widehat{B}_j(p)\| = \mathcal{O}_P(1), \tag{3.26}$$
- $$E^*\Big(\underline{\varepsilon}_t^*(j)^{2w}\Big) \xrightarrow{P} E\Big(\underline{\varepsilon}_1(j)^{2w}\Big) \quad \forall\, w \leq h+2,\ \forall\, j = 1, \ldots, q, \tag{3.27}$$
- $$\underline{\varepsilon}_t^* \xrightarrow{d^*} \underline{\varepsilon}_t \ \ \textit{in } P\textit{-prob.}\ \ \forall\, t \in \mathbb{Z}, \tag{3.28}$$
- $$(\underline{X}_{t_1}^*, \ldots, \underline{X}_{t_d}^*) \xrightarrow{d^*} (\underline{\widetilde{X}}_{t_1}, \ldots, \underline{\widetilde{X}}_{t_d}) \ \ \textit{in } P\textit{-prob.}\ \ \forall\, d \in \mathbb{N},\ \forall\, t_1, \ldots, t_d \in \mathbb{Z}. \tag{3.29}$$

The proof can be found in section 3.8.

3.5 Applications

In the following, we will discuss the check criterion developed in the previous section by means of some important statistics, namely the sample mean, sample autocovariances and sample autocorrelations.

Example 3.12. (Sample mean) Let $(\underline{X}_t)_{t\in\mathbb{Z}}$ be a strictly stationary process with mean $\underline{\mu}$ such that the centered version $(\underline{X}_t - \underline{\mu})$ fulfils Assumption 5. Theorem 3.10 then applies to $(\underline{X}_t - \underline{\mu})$ and we can consider $T_n = n^{-1}\sum_{t=1}^{n}(\underline{X}_t - \underline{\mu})$ which falls into the class of statistics defined by Assumption 7. For the sample mean $\underline{\overline{X}}_n = n^{-1}\sum_{t=1}^{n}\underline{X}_t$ it is known under standard regularity conditions that

$$\sqrt{n}\, T_n = \sqrt{n}(\underline{\overline{X}}_n - \underline{\mu}) \xrightarrow{d} \mathcal{N}\Big(\underline{0}\,,\, \textstyle\sum_{h\in\mathbb{Z}} \Gamma_{\underline{X}}(h)\Big),$$

where $\Gamma_{\underline{X}}(\cdot)$ denotes the autovariance function of $(\underline{X}_t)$. For the companion process we get analogously

$$\sqrt{n}\, \widetilde{T}_n = \sqrt{n}\Big(\underline{\overline{\widetilde{X}}}_n - \underline{\mu}\Big) \xrightarrow{d} \mathcal{N}\Big(\underline{0}\,,\, \textstyle\sum_{h\in\mathbb{Z}} \Gamma_{\underline{\widetilde{X}}}(h)\Big),$$

where $\Gamma_{\widetilde{\underline{X}}}(\cdot)$ denotes the autocovariance function of the companion process $(\widetilde{\underline{X}_t})$. Since it is easy to see from the definitions of $(\underline{X}_t)$ and $(\widetilde{\underline{X}_t})$ that it holds $\Gamma_{\widetilde{\underline{X}}}(h) = \Gamma_{\underline{X}}(h)$ for all $h \in \mathbb{Z}$, the limiting distributions of $\sqrt{n}\,T_n$ and $\sqrt{n}\,\widetilde{T}_n$ coincide, and Theorem 3.10 shows that the VAR sieve bootstrap procedure works asymptotically under very mild conditions.

Example 3.13. (Sample autocovariances) For the sample autocovariance $T_n = \widehat{\Gamma}(h)$, as defined in (3.17), Kreiss, Paparoditis and Politis (2011) show in the univariate setting $q = 1$ that the AR sieve bootstrap procedure fails in general. Of course, the same holds true in the more general case of multivariate processes. In particular, their Example 3.2 also discusses the case of data stemming from a univariate *linear* process, where the procedure may still fail. This result, although against intuition at first glance, emphasizes the fact that Theorem 3.1 in Kreiss, Paparoditis and Politis (2011), as well as Theorem 3.10 in the present thesis, do not imply that the sieve bootstrap procedure automatically works as soon as the data are generated by a linear process. Only by comparing the limiting distributions of a particular statistic for both the underlying and the companion process, one can decide whether the procedure works asymptotically or not.

Example 3.14. (Sample autocorrelations) Consider a linear process $(\underline{X}_t)_{t\in\mathbb{Z}}$, i.e. $\underline{X}_t = \sum_{j=-\infty}^{\infty} B_j\,\underline{\varepsilon}_{t-j}$ with $(\underline{\varepsilon}_t)$ i.i.d. and finite fourth moments, and its cross-correlation function $\rho_{i,k}(h) = \Gamma^{(i,k)}(h)/(\Gamma^{(i,i)}(0)\Gamma^{(k,k)}(0))^{1/2}$. Under standard assumptions it holds

$$\sqrt{n}\,(\widehat{\rho}_{i,k}(h) - \rho_{i,k}(h)) \xrightarrow{d} \mathcal{N}(0, \tau^2_{\underline{X}}),$$

where it is known that $\tau^2_{\underline{X}}$ can not be expressed exclusively in terms of second order quantities of $(\underline{X}_t)$. Therefore, with the same reasoning as in Example 3.2 in Kreiss, Paparoditis and Politis (2011), the VAR sieve bootstrap fails in general, although the data are generated by a linear process. However, there are three special cases of linear processes for which the procedure works asymptotically:

(a) In the case of univariate linear processes the variance τ^2_X is given by Bartlett's formula, cf. Brockwell and Davis (1991), Theorem 7.2.1, and depends only on the autocorrelations $\rho_X(j)$ of the underlying process. Since it holds $\rho_{\widetilde{X}}(j) = \rho_X(j)$ for all $j \in \mathbb{Z}$, the limiting variances τ^2_X and $\tau^2_{\widetilde{X}}$ of T_n and $\widetilde{T}_n$ coincide and the bootstrap procedure works. This is Example 3.3 in Kreiss, Paparoditis and Politis (2011).

(b) If the coefficient matrices B_j are diagonal matrices and the component processes $(\varepsilon_t(j))_{t\in\mathbb{Z}}$, $j = 1,\ldots,q$, are independent of each other, then Theorem 11.2.2 from Brockwell and Davis (1991) yields

$$\sqrt{n}\,\widehat{\rho}_{i,k}(h) \xrightarrow{d} \mathcal{N}\Big(0, \sum\nolimits_{r\in\mathbb{Z}} \rho_{i,i}(r)\rho_{k,k}(r)\Big).$$

Again, since all cross-correlations are identical for $(\underline{X}_t)$ and $(\widetilde{\underline{X}}_t)$, Theorem 3.10 proves asymptotic validity of the VAR sieve bootstrap.

(c) If $(\underline{X}_t)$ is Gaussian, Theorem 11.2.3 in Brockwell and Davis (1991) implies

$$\begin{aligned}\tau^2_{\underline{X}} = \sum_{j\in\mathbb{Z}}\Big[&\rho_{i,i}(j)\rho_{k,k}(j) + \rho_{i,k}(j+h)\rho_{k,i}(j-h)\\ &- (\rho_{i,k}(h) + \rho_{i,k}(j))\Big(\rho_{i,i}(j)\rho_{k,i}(j+h) + \rho_{k,k}(j)\rho_{k,i}(j-h)\Big)\\ &+ \rho_{i,k}(h)^2\Big(\frac{1}{2}\rho_{i,i}(j)^2 + \rho_{i,k}(j)^2 + \frac{1}{2}\rho_{k,k}(j)^2\Big)\Big].\end{aligned}$$

Since $\tau^2_{\underline{X}}$ depends only on the cross-correlations of $(\underline{X}_t)$, Theorem 3.10 shows that the VAR sieve bootstrap procedure is asymptotically valid.

3.6 Nonparametric trend estimation

The VAR sieve bootstrap procedure may also be used to approximate the distribution of trend function estimators in multivariate time series data. Consider the deterministic trend function $\underline{m} : [0,1] \to \mathbb{R}^q$ and observations $\underline{Y}_{1,n},\ldots,\underline{Y}_{n,n}$ fulfilling

$$\underline{Y}_{t,n} = \underline{m}(t/n) + \underline{X}_t, \quad t = 1,\ldots,n\,, \tag{3.30}$$

where $(\underline{X}_t)_{t\in\mathbb{Z}}$ is a strictly stationary process with mean zero which fulfils Assumption 5 from section 3.2. The trend funtion may be estimated by the kernel estimator

$$\widehat{\underline{m}}(x) = \frac{1}{n\delta}\sum_{t=1}^{n} K\Big(\frac{x - t/n}{\delta}\Big)\,\underline{Y}_{t,n}, \tag{3.31}$$

where $\delta = \delta(n)$ is the bandwidth. We impose the following conditions on the bandwidth and the kernel:

Assumption 9. *The bandwidth fulfils $\delta(n) \to 0$ and $n\,\delta(n) \to \infty$, as $n \to \infty$. The kernel $K : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is bounded and has support $[-1,1]$. Furthermore, K is Lipschitz and it holds $\int_{-1}^{1} K^2(x)\,dx < \infty$.*

It is worth mentioning that we could as well use different kernels for each component of $\underline{\widehat{m}}$. The results would still be valid while the notation would become considerably more complex, therefore we use the same kernel for all components.

The goal is to approximate the distribution of $(n\delta)^{1/2}(\underline{\widehat{m}}(x) - E(\underline{\widehat{m}}(x)))$. Starting with a given sample of observations $\underline{Y}_{1,n}, \ldots, \underline{Y}_{n,n}$, we propose to invoke the VAR sieve bootstrap procedure in the following way:

(a) Calculate

$$\underline{\widehat{X}}_t := \underline{Y}_{t,n} - \underline{\widehat{m}}(t/n), \qquad t = 1, \ldots, n,$$

where $\underline{\widehat{m}}(\cdot)$ is given by (3.31).

(b) Apply step 1 and 2 of the VAR sieve bootstrap procedure, as described in section 3.2, to $(\underline{\widehat{X}}_1, \ldots, \underline{\widehat{X}}_n)$. This yields a bootstrap sample $(\underline{X}_1^*, \ldots, \underline{X}_n^*)$.

(c) Calculate bootstrap observations $\underline{Y}_{1,n}^*, \ldots, \underline{Y}_{n,n}^*$ according to

$$\underline{Y}_{t,n}^* := \underline{\widehat{m}}(t/n) + \underline{X}_t^*, \qquad t = 1, \ldots, n.$$

(d) Calculate the bootstrap version of the trend function estimator

$$\underline{\widehat{m}}^*(x) = \frac{1}{n\delta} \sum_{t=1}^{n} K\Big(\frac{x - t/n}{\delta}\Big) \underline{Y}_{t,n}^*$$

and denote the result by $\underline{\widehat{m}}_{(1)}^*(x)$.

(e) Repeat steps $(b) - (d)$ M times to obtain independent realisations $\underline{\widehat{m}}_{(1)}^*(x), \ldots,$ $\underline{\widehat{m}}_{(M)}^*(x)$, where M is sufficiently large. Use the empirical distribution of $\mathcal{L}\Big((n\delta)^{1/2}(\underline{\widehat{m}}^*(x) - E^*(\underline{\widehat{m}}^*(x)))\Big)$, based on the realisations $\underline{\widehat{m}}_{(1)}^*(x), \ldots, \underline{\widehat{m}}_{(M)}^*(x)$, as an approximation for $\mathcal{L}\Big((n\delta)^{1/2}(\underline{\widehat{m}}(x) - E(\underline{\widehat{m}}(x)))\Big)$.

Here, E^* means expectation conditional on $\underline{Y}_{1,n}, \ldots, \underline{Y}_{n,n}$. In order to prove asymptotic validity of this procedure, we require the following assumption:

Assumption 10. *For each pair $r, s \in \{1, \ldots, q\}$, we have $E(\varepsilon_1(r)^2) < \infty$ as well as*

$$\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t(r)\varepsilon_t(s) \xrightarrow{P} E\Big(\varepsilon_1(r)\varepsilon_1(s)\Big),$$

where $\varepsilon_t(1), \ldots, \varepsilon_t(q)$ are the components of $\underline{\varepsilon}_t$.

We can now state the following result about the asymptotic distribution of the bootstrap version of the trend estimator:

Theorem 3.15. *Let $\underline{Y}_{1,n}, \ldots, \underline{Y}_{n,n}$ be generated by (3.30) and $(\underline{X}_t)_{t\in\mathbb{Z}}$ fulfil Assumption 5 for $r = 1$. Assume Assumptions 6, 9 and 10 hold. Then we have*

$$\mathcal{L}\Big((n\delta)^{1/2}(\underline{\widehat{m}}^*(x) - E^*(\underline{\widehat{m}}^*(x)))\,|\,\underline{Y}_{1,n}, \ldots, \underline{Y}_{n,n}\Big) \stackrel{w}{\longrightarrow} \mathcal{N}(\underline{0}, \Sigma)$$

in probability as $n \to \infty$, where $\Sigma = \int_{-1}^{1} K^2(u)\,du \cdot \sum_{h\in\mathbb{Z}} \Gamma_{\underline{X}}(h)$.

For the non-bootstrap quantities it is known that the following result holds under standard regularity conditions:

$$(n\delta)^{1/2}(\underline{\widehat{m}}(x) - E(\underline{\widehat{m}}(x))) \stackrel{d}{\longrightarrow} \mathcal{N}(\underline{0}, \Sigma),$$

where Σ is the limiting variance given in Theorem 3.15, cf. for example Altman (1990) and Hart (1991) for the univariate case, which carries over to the multivariate setting presented here (alternatively, to obtain the multivariate result, one can follow along the lines of the proof of Theorem 3.15, merely replacing bootstrap quantities by their theoretical counterparts). Therefore, Theorem 3.15 shows that the VAR sieve bootstrap procedure is asymptotically valid for the non-parametric trend estimation introduced in this section.

The proof of Theorem 3.15 can be found in section 3.7. It relies, among other things, on the following assertions:

Lemma 3.16. *Under the assumptions of Theorem 3.15 the following assertions hold true for $r, s \in \{1, \ldots, q\}$ and for all $h \in \mathbb{Z}$. The expressions $\underline{X}^*_{t,M}$ and $\underline{L}^*_{n,M}(x)$ are defined in the proof of Theorem 3.15, cf. section 3.7.*

- $E^*(\varepsilon^*_0(r)\,\varepsilon^*_0(s)) = E(\varepsilon_0(r)\,\varepsilon_0(s)) + o_P(1),$ (3.32)
- $\mathrm{Cov}^*(\underline{X}^*_{0,M}(r), \underline{X}^*_{h,M}(s)) = \mathrm{Cov}(\underline{X}_{0,M}(r), \underline{X}_{h,M}(s)) + o_P(1),$ (3.33)
- $E^*(\underline{L}^*_{n,M}(x)^{(r)} \cdot \underline{L}^*_{n,M}(x)^{(s)}) = \Sigma^{(r,s)}_M + o_P(1),$ (3.34)
- $\underline{c}^T \Sigma_M\, \underline{c} \longrightarrow \underline{c}^T \Sigma\, \underline{c}, \qquad$ *as* $M \to \infty,$ (3.35)

where $\Sigma^{(r,s)}_M = \int_{-1}^{1} K^2(u)\,du \cdot \sum_{|h|\le M} \mathrm{Cov}[\underline{X}_{0,M}(r), \underline{X}_{h,M}(s)]$.

The proof can be found in Section 3.8.

3.7 Proofs of the main results

3.7.1 Proof of Theorem 3.10

Theorem 3.10 is a direct generalization of Theorem 3.3 in Bühlmann (1995) to the case of multivariate processes. We will therefore omit the lengthy proof here, and only show how the critical points carry over to our setting. In the following we will use Bühlmann's notation as far as possible. However, it should be noted that we prove validity for the companion process which is denoted by $(\underline{\widetilde{X}}_t)$ while Bühlmann proves validity of his bootstrap scheme for a process denoted by (X_t).
The companion process and the bootstrap process possess the representations (3.24) and (3.22), respectively, and we denote the truncated versions by

$$\underline{\widetilde{X}}_{t,M} = \sum_{k=1}^{M} B_k \, \underline{\widetilde{\varepsilon}}_{t-k} + \underline{\widetilde{\varepsilon}}_t \quad \text{and} \quad \underline{X}^*_{t,M} = \sum_{k=1}^{M} \widehat{B}_k(p) \, \underline{\varepsilon}^*_{t-k} + \underline{\varepsilon}^*_t. \tag{3.36}$$

In analogy to Bühlmann's notation we denote $\mathbf{X}_t := \text{vec}(\underline{\widetilde{X}}_t, \ldots, \underline{\widetilde{X}}_{t+m-1})$, $\mathbf{X}_{t,M} := \text{vec}(\underline{\widetilde{X}}_{t,M}, \ldots, \underline{\widetilde{X}}_{t+m-1,M})$ and $\mathbf{X}^*_t$, $\mathbf{X}^*_{t,M}$ analogously. Since the multivariate process $(\underline{X}^*_{t,M})$ is strictly stationary and M-dependent, the process $\mathbf{X}^*_{t,M}$ is strictly stationary and $(M+m-1)$-dependent. Hence, we can show Bühlmann's assertion (5.27) exactly along the lines using auxiliary result (3.29) from Lemma 3.11 instead of Bühlmann's Corollary 5.6.
Next, we define for arbitrary but fixed $\underline{c} \in \mathbb{R}^d$ the function $l(\underline{x}) := \sum_{i=1}^{d} c_i \, g_i(\underline{x})$. Using the same blocking technique as Bühlmann we can prove the CLT for the truncated process, i.e.

$$(n-m+1)^{-1/2} \sum_{t=1}^{n-m+1} \Big(g(\mathbf{X}^*_{t,M}) - E^*[g(\mathbf{X}^*_{t,M})] \Big) \xrightarrow{d^*} \mathcal{N}(0, \Sigma_M) \;\text{ in } P\text{-prob.},$$

along the lines, if we can ensure that

$$E^* |l(\mathbf{X}^*_{t,M})|^{2+2/(h+1)} = \mathcal{O}_P(1) \tag{3.37}$$

holds in our setting. We will do this using the following Taylor expansion:

$$l(\mathbf{X}^*_{t,M}) = \sum_{|\alpha| \le h-1} \frac{1}{\alpha!} \, D^\alpha \, l(\underline{0}) \cdot (\mathbf{X}^*_{t,M})^\alpha + \sum_{|\alpha|=h} \frac{1}{\alpha!} \, D^\alpha \, l(\underline{\tau}) \cdot (\mathbf{X}^*_{t,M})^\alpha, \tag{3.38}$$

for some $\underline{\tau} = \rho \cdot \mathbf{X}^*_{t,M}$, $\rho \in [0,1]$, where $\alpha \in \mathbb{N}_0^{mq}$, $|\alpha| = \alpha_1 + \ldots + \alpha_{mq}$, $(\underline{x})^\alpha = x_1^{\alpha_1} \cdot \ldots \cdot x_{mq}^{\alpha_{mq}}$ and

$$D^\alpha \, l(\underline{x}) := \frac{\partial^{|\alpha|} \, l(\underline{y})}{\partial y_1^{\alpha_1} \ldots \partial y_{mq}^{\alpha_{mq}}} \Big|_{\underline{y}=\underline{x}}.$$

Now, let $\| \cdot \|_{*r}$ denote the usual $\mathcal{L}_r$-norm with respect to P^*. In particular, let $r := (2 + 2/(h+1))\,|\alpha|$ for any $|\alpha| \leq h-1$. Consider $\mathbf{X}^*_{t,M}(i)$ for $i \in \{1, \ldots, q\}$, i.e. one of the first q components of $\mathbf{X}^*_{t,M}$. From (3.36) we have

$$\mathbf{X}^*_{t,M}(i) = \sum_{k=1}^{M} \sum_{j=1}^{q} \widehat{B}_k(p)^{(i,j)}\, \underline{\varepsilon}^*_{t-k}(j) + \underline{\varepsilon}^*_t(i).$$

With the strict stationarity of $\underline{\varepsilon}^*_t$ and (3.3) we get

$$\begin{aligned} \| \mathbf{X}^*_{t,M}(i) \|_{*r} &\leq \sum_{j=1}^{q} \| \underline{\varepsilon}^*_t(j) \|_{*r} \sum_{k=1}^{M} \left| \widehat{B}_k(p)^{(i,j)} \right| + \| \underline{\varepsilon}^*_t(i) \|_{*r} \\ &\leq \sum_{j=1}^{q} \| \underline{\varepsilon}^*_t(j) \|_{*r} \sum_{k=1}^{\infty} \left\| \widehat{B}_k(p) \right\| + \| \underline{\varepsilon}^*_t(i) \|_{*r} \\ &= \mathcal{O}_P(1) \end{aligned} \tag{3.39}$$

for p large enough because of (3.26) and

$$\Big(\| \underline{\varepsilon}^*_t(j) \|_{*r} \Big)^{(2+2/(h+1))\,|\alpha|} = E^* |\underline{\varepsilon}^*_t(j)|^{(2+2/(h+1))\,|\alpha|},$$

which converges in probability to a finite limit, cf. (3.27) and Assumption 8 (note that $(2 + 2/(h+1))\,|\alpha| \leq 2(h+2)$ since $|\alpha| \leq h-1$). The same holds true for all other components of $\mathbf{X}^*_{t,M}$ due to analogous arguments.
From expansion (3.38) we get for $s = 2 + 2/(h+1)$

$$\|l(\mathbf{X}^*_{t,M})\|_{*s} \leq \sum_{|\alpha| \leq h-1} \frac{|D^\alpha\, l(\underline{0})|}{\alpha!} \cdot \left\| (\mathbf{X}^*_{t,M})^\alpha \right\|_{*s} + \sum_{|\alpha|=h} \frac{1}{\alpha!} \left\| D^\alpha\, l(\underline{\tau}) \cdot (\mathbf{X}^*_{t,M})^\alpha \right\|_{*s}. \tag{3.40}$$

The first summand on the right-hand side of (3.40) is $\mathcal{O}_P(1)$ because of (3.39) and since

$$\left\| (\mathbf{X}^*_{t,M})^\alpha \right\|_{*s} \leq \left\| \mathbf{X}^*_{t,M}(1) \right\|_{*r}^{\alpha_1} \cdot \ldots \cdot \left\| \mathbf{X}^*_{t,M}(mq) \right\|_{*r}^{\alpha_{mq}}$$

follows from Hölder's inequality. Since the h-th derivative of l is Lipschitz due to Assumption 7, we get

$$\begin{aligned} & \left\| D^\alpha\, l(\underline{\tau}) \cdot (\mathbf{X}^*_{t,M})^\alpha \right\|_{*s} \\ \leq\ & \left| D^\alpha\, l(\underline{0}) \right| \cdot \left\| (\mathbf{X}^*_{t,M})^\alpha \right\|_{*s} + \left\| \, |D^\alpha\, l(\underline{\tau}) - D^\alpha\, l(\underline{0})| \cdot (\mathbf{X}^*_{t,M})^\alpha \right\|_{*s} \\ \leq\ & \left| D^\alpha\, l(\underline{0}) \right| \cdot \left\| (\mathbf{X}^*_{t,M})^\alpha \right\|_{*s} + \left\| L(|\tau_1| + \ldots + |\tau_{mq}|) \cdot (\mathbf{X}^*_{t,M})^\alpha \right\|_{*s} \\ \leq\ & \left| D^\alpha\, l(\underline{0}) \right| \cdot \left\| (\mathbf{X}^*_{t,M})^\alpha \right\|_{*s} + L\, \rho \sum_{j=1}^{mq} \left\| \mathbf{X}^*_{t,M}(j) \cdot (\mathbf{X}^*_{t,M})^\alpha \right\|_{*s} \\ =\ & \mathcal{O}_P(1), \end{aligned}$$

using the same arguments as before. Hence, (3.40) yields $\|l(\mathbf{X}^*_{t,M})\|_{*s} = \mathcal{O}_P(1)$ which is (3.37).

We can now derive Bühlmann's assertion (5.39), i.e. $(\Sigma_M)_{u,v} \to (\Sigma)_{u,v}$ as $M \to \infty$ for all $u, v \in \{1, \dots, mq\}$, along the lines of his proof with straightforward Taylor expansions as in (3.38) above. Here, we need $\sum_{j=1}^{\infty} j\,\|B_j\| < \infty$ which is provided by (3.9).

As the final part of the proof we establish Bühlmann's assertion (5.40) in our setting. Defining for all $k > m$

$$\begin{aligned}\tilde{\mathbf{X}}^*_k &:= \operatorname{vec}\big(\textstyle\sum_{j=1}^{k-m} \widehat{B}_j(p)\,\underline{\varepsilon}^*_{k-j} + \underline{\varepsilon}^*_k, \dots, \sum_{j=1}^{k-m} \widehat{B}_j(p)\,\underline{\varepsilon}^*_{k+m-j} + \underline{\varepsilon}^*_{k+m}\big),\\ \tilde{\mathbf{X}}^*_{k,M} &:= \operatorname{vec}\big(\textstyle\sum_{j=1}^{M\wedge(k-m)} \widehat{B}_j(p)\,\underline{\varepsilon}^*_{k-j} + \underline{\varepsilon}^*_k, \dots, \sum_{j=1}^{M\wedge(k-m)} \widehat{B}_j(p)\,\underline{\varepsilon}^*_{k+m-j} + \underline{\varepsilon}^*_{k+m}\big),\end{aligned}$$

one can again follow along the lines of Bühlmann (1995), adapting the Taylor expansions to the multivariate setting as shown above and using (3.26), which completes the proof. $\square$

3.7.2 Proof of Theorem 3.15

Defining the abbreviation $\underline{L}^*_n(x) := (n\delta)^{1/2}(\underline{\widehat{m}}^*(x) - E^*(\underline{\widehat{m}}^*(x)))$, the goal in this proof is to show

$$\underline{L}^*_n(x) \xrightarrow{d^*} \mathcal{N}(\underline{0}, \Sigma) \quad \text{in prob.},$$

where Σ is given by Theorem 3.15. By the Cramér-Wold device, this follows if we can show

$$\underline{c}^T\,\underline{L}^*_n(x) \xrightarrow{d^*} \mathcal{N}(0, \underline{c}^T\Sigma\,\underline{c}) \quad \text{in prob.},$$

for arbitrary $\underline{c} \neq \underline{0}$. Since we assume that $(\underline{X}_t)_{t\in\mathbb{Z}}$ fulfils Assumption 5 it possesses the moving average representation (3.5) while the bootstrap process is given by (3.22). For arbitrary $M > 0$ we define the truncated versions by

$$\underline{X}_{t,M} = \sum_{k=0}^{M} B_k\,\underline{\varepsilon}_{t-k} \quad \text{and} \quad \underline{X}^*_{t,M} = \sum_{k=0}^{M} \widehat{B}_k(p)\,\underline{\varepsilon}^*_{t-k},$$

where $B_0 = \widehat{B}_0(p) = I$. In analogy to step (3) of the bootstrap procedure above, we define $\underline{Y}^*_{t,M} := \underline{\widehat{m}}(t/n) + \underline{X}^*_{t,M}$ for $t = 1, \dots, n$. Replacing $\underline{Y}^*_{t,n}$ with $\underline{Y}^*_{t,M}$ in the definition of $\underline{\widehat{m}}^*(x)$ gives $\underline{\widehat{m}}^*_M(x)$. Subsequently replacing $\underline{\widehat{m}}^*(x)$ with its truncated version in the definition of $\underline{L}^*_n(x)$ yields $\underline{L}^*_{n,M}(x)$.

Let $\underline{c} \neq \underline{0}$ be arbitrary. The strategy is to first prove

$$\underline{c}^T \underline{L}^*_{n,M}(x) \xrightarrow{d^*} \mathcal{N}(0, \underline{c}^T \Sigma_M \, \underline{c}) \quad \text{in prob.,}$$

where $\Sigma_M^{(r,s)} = \int_{-1}^{1} K^2(u)\, du \cdot \sum_{h=-M}^{M} \mathrm{Cov}[\underline{X}_{0,M}(r), \underline{X}_{h,M}(s)]$. In order to derive the limiting variance of $\underline{c}^T \underline{L}^*_{n,M}(x)$ we use (3.34) from Lemma 3.16 to obtain

$$\mathrm{Var}^*(\underline{c}^T \underline{L}^*_{n,M}(x)) = \underline{c}^T E^*(\underline{L}^*_{n,M}(x) \cdot \underline{L}^*_{n,M}(x)^T)\, \underline{c} \ \xrightarrow{P}\ \underline{c}^T \Sigma_M \, \underline{c}. \tag{3.41}$$

For this limiting variance we have already established

$$\underline{c}^T \Sigma_M \, \underline{c} \longrightarrow \underline{c}^T \Sigma \, \underline{c}, \qquad \text{as } M \to \infty,$$

cf. (3.35) from Lemma 3.16.
We will now show

$$(1/v^*)\, \underline{c}^T \underline{L}^*_{n,M}(x) \xrightarrow{d^*} \mathcal{N}(0,1) \quad \text{in } P\text{-probability} \tag{3.42}$$

for all $\underline{c} \neq \underline{0}$, where $v^* := (\mathrm{Var}^*(\underline{c}^T \underline{L}^*_{n,M}(x)))^{1/2}$. Since all eigenvalues of Σ are bounded away from zero per Assumption 5, Σ is positive definite, i.e. $\underline{c}^T \Sigma \, \underline{c} > 0$. Without loss of generality we only consider M large enough such that $\underline{c}^T \Sigma_M \, \underline{c} > 0$ also, which is possible because of (3.35). Hence, we immediately get from (3.41) that

$$1/v^* = \mathcal{O}_P(1). \tag{3.43}$$

We will now introduce sequences of positive integers $N(n), a(n), b(n)$ with certain asymptotic properties. These sequences will be used to decompose the expression in (3.42). Assume without loss of generality that $n\delta \in \mathbb{N}$. We choose $N(n)$ such that $N(n) \to +\infty$ as $n \to \infty$ but at a slow rate, namely $N(n) = o((n\delta)^{1/4})$ and such that $2n\delta/N(n) \in \mathbb{N}$ for all $n \in \mathbb{N}$. Then we decompose the increasing sequence of positive integers $2n\delta/N(n) = a(n) + b(n)$, where $a(n), b(n) \in \mathbb{N}$ for all n, $b(n) \to +\infty$ as $n \to \infty$ and $b(n) = o((n\delta)^{1/4})$. For example, $b(n)$ could be chosen as $(n\delta)^{1/5}$ (rounded to integer values) in order to fulfil these properties. It follows immediately

$$\frac{b(n)}{a(n)} = \frac{b(n)}{2n\delta/N(n) - b(n)} = (2n\delta)^{-1} \frac{b(n)N(n)}{1 - b(n)N(n)/2n\delta} = o(1), \tag{3.44}$$

i.e. $a(n)$ dominates $b(n)$. Using the definition of $\underline{L}^*_{n,M}(x)$ one can easily derive for the expression in (3.42) that

$$(1/v^*)\, \underline{c}^T \underline{L}^*_{n,M}(x) = (1/v^*)(n\delta)^{-1/2} \sum_{t=1}^{n} K\Big(\frac{x - t/n}{\delta}\Big) \underline{c}^T \underline{X}^*_{t,M} \tag{3.45}$$

Since K is zero outside of $[-1,1]$ only $2n\delta$ consecutive summands of the n summands on the right-hand side of (3.45) are non-zero (remember the assumption $n\delta \in \mathbb{N}$). Let $t_n \in \{1,\dots,n\}$ be such that the summands associated with indices $t = t_n + 1,\dots,t_n{+}2n\delta$ in (3.45) are exactly the non-zero ones. Using the sequences introduced above we can now decompose $(1/v^*)\,\underline{c}^T\underline{L}^*_{n,M}(x)$ via (3.45) as

$$(1/v^*)\,\underline{c}^T\underline{L}^*_{n,M}(x) = (1/v^*)(n\delta)^{-1/2}\sum_{i=1}^{N(n)} A_{n,i} + (1/v^*)(n\delta)^{-1/2}\sum_{i=1}^{N(n)} B_{n,i}, \qquad (3.46)$$

where (suppressing the dependence of a and b on n)

$$A_{n,i} := \sum_{t=t_n+1+(i-1)(a+b)}^{t_n+ia+(i-1)b} K\Big(\frac{x-t/n}{\delta}\Big)\underline{c}^T\underline{X}^*_{t,M}, \quad i=1,\dots,N(n)\,,$$
$$B_{n,i} := \sum_{t=t_n+1+ia+(i-1)b}^{t_n+i(a+b)} K\Big(\frac{x-t/n}{\delta}\Big)\underline{c}^T\underline{X}^*_{t,M}, \quad i=1,\dots,N(n).$$

Note that each $A_{n,i}$ consists of $a(n)$ summands, each $B_{n,i}$ of $b(n)$ ones. The strategy is to show that the second summand on the right-hand side of (3.46) is $o_{P^*}(1)$ in P-probability, while the first one converges in distribution to $\mathcal{N}(0,1)$ in P-probability which will prove the desired assertion (3.42).

Firstly, the strict stationarity of $(\underline{\varepsilon}^*_t)$ combined with (3.27) ensures that $E^*|\varepsilon^*_t(j)| = \mathcal{O}_P(1)$ uniformly for all $t\in\mathbb{Z}$ and all $j=1,\dots,q$. Hence,

$$E^*|\underline{X}^*_{t,M}(j)| \le \sum_{k=0}^{M}\sum_{l=1}^{q} |\widehat{B}_k(p)^{(j,l)}|\, E^*|\varepsilon^*_{t-k}(l)| \le q\sum_{k=0}^{M}\|\widehat{B}_k(p)\|\cdot\mathcal{O}_P(1) = \mathcal{O}_P(1)$$

uniformly for all t because of (3.26). This implies $E^*|\underline{c}^T\underline{X}^*_{t,M}| = \mathcal{O}_P(1)$ uniformly and therefore

$$E^*|B_{n,i}| \le \sum_{t=t_n+1+ia+(i-1)b}^{t_n+i(a+b)} \Big|K\Big(\frac{x-t/n}{\delta}\Big)\Big| E^*|\underline{c}^T\underline{X}^*_{t,M}| = b(n)\cdot\mathcal{O}_P(1)$$

for each $i=1,\dots,N(n)$. We can now use Markov's inequality and (3.43) to derive

$$\begin{aligned} P^*\Big(\Big|(1/v^*)(n\delta)^{-1/2}\sum_{i=1}^{N(n)} B_{n,i}\Big| > \kappa\Big) &\le (1/\kappa v^*)(n\delta)^{-1/2}\sum_{i=1}^{N(n)} E^*|B_{n,i}| \\ &\le (1/\kappa v^*)(n\delta)^{-1/2}\,N(n)\,b(n)\cdot\mathcal{O}_P(1) \\ &= (n\delta)^{-1/2}\,o((n\delta)^{1/2})\cdot\mathcal{O}_P(1) \\ &= o_P(1), \end{aligned}$$

or in other words

$$(1/v^*)(n\delta)^{-1/2}\sum_{i=1}^{N(n)} B_{n,i} \xrightarrow{P^*} 0 \quad \text{in } P\text{-probability.} \tag{3.47}$$

We will now show for the first summand in (3.46) that

$$(1/v^*)(n\delta)^{-1/2}\sum_{i=1}^{N(n)} A_{n,i} \xrightarrow{d^*} \mathcal{N}(0,1) \quad \text{in } P\text{-probability,} \tag{3.48}$$

as $n \to \infty$. Define $\tau_n^2 := \mathrm{Var}^*\big(\sum_{i=1}^{N(n)} A_{n,i}\big)$. Observe that, for all n large enough such that $b(n) > M$, the $A_{n,1},\ldots,A_{n,N(n)}$ are (conditionally) independent random variables. Therefore, if we can show that the Lindeberg condition

$$\tau_n^{-2}\sum_{i=1}^{N(n)} E^*(A_{n,i}^2 \cdot \mathbb{1}\{|A_{n,i}| > \varepsilon\,\tau_n\}) = o_P(1) \quad \forall\, \varepsilon > 0 \tag{3.49}$$

holds, it follows

$$\tau_n^{-1}\sum_{i=1}^{N(n)} A_{n,i} \xrightarrow{d^*} \mathcal{N}(0,1) \quad \text{in } P\text{-probability.} \tag{3.50}$$

That this is equivalent to (3.48) can be obtained from

$$(\tau_n/v^*)(n\delta)^{-1/2} \xrightarrow{P} 1, \tag{3.51}$$

using Slutsky's Theorem. We now prove (3.51). From the definitions of τ_n^2 and v^* we get

$$\begin{aligned}
&\Big((\tau_n/v^*)(n\delta)^{-1/2}\Big)^2 = \mathrm{Var}^*\Big((1/v^*)(n\delta)^{-1/2}\sum_{i=1}^{N(n)} A_{n,i}\Big)\\
&= \mathrm{Var}^*\Big((1/v^*)\underline{c}^T\underline{L}_{n,M}^*(x) - (1/v^*)(n\delta)^{-1/2}\sum_{i=1}^{N(n)} B_{n,i}\Big)\\
&= 1 + \mathrm{Var}^*\Big((1/v^*)(n\delta)^{-1/2}\sum_{i=1}^{N(n)} B_{n,i}\Big) - \frac{2}{(v^*)^2}\cdot \mathrm{Cov}^*\Big(\underline{c}^T\underline{L}_{n,M}^*(x), (n\delta)^{-1/2}\sum_{i=1}^{N(n)} B_{n,i}\Big)
\end{aligned}$$

The second summand on the right-hand side converges to zero in P-probability because of (3.47), while the absolute value of the third summand can be bounded by

$$2\cdot\Big(\mathrm{Var}^*\Big((1/v^*)\underline{c}^T\underline{L}_{n,M}^*(x)\Big)\Big)^{1/2}\cdot\Big(\mathrm{Var}^*\Big((1/v^*)(n\delta)^{-1/2}\sum_{i=1}^{N(n)} B_{n,i}\Big)\Big)^{1/2}$$

$$= 2 \cdot \Big(\mathrm{Var}^*\Big((1/v^*)(n\delta)^{-1/2}\sum_{i=1}^{N(n)} B_{n,i}\Big)\Big)^{1/2},$$

which converges to zero in P-probability by the same argument. This yields (3.51). We now establish a moment condition on the $A_{n,i}$. Let $\eta > 0$ be such that it holds $E\,|\varepsilon_t(k)|^{2+\eta} < \infty$ and define $Z_t^* := K\Big((x - t/n)/\delta\Big)\,\underline{c}^T\underline{X}_{t,M}^*$. Above, we have already shown $E^*|Z_t^*| = \mathcal{O}_P(1)$. On the same way we can easily derive $E^*(|Z_t^*|^{2+\eta}) = \mathcal{O}_P(1)$ uniformly for all $t \in \mathbb{Z}$. Since the Z_t^* are (conditionally) M-dependent random variables with mean zero we can apply Corollary A.1 in Romano and Wolf (2000) to $A_{n,i}$ and get

$$E^*(|A_{n,i}|^{2+\eta}) = \mathcal{O}_P\Big(a(n)^{1+\eta/2}\Big) \quad \text{uniformly } \forall\, i = 1,\ldots,N(n). \tag{3.52}$$

Now, we are able to prove Lindeberg condition (3.49). First we obtain from Hölder's and Markov's inequalities

$$\begin{aligned}
&\tau_n^{-2}\sum_{i=1}^{N(n)} E^*(A_{n,i}^2 \cdot \mathbb{1}\{|A_{n,i}| > \varepsilon\,\tau_n\})\\
&\le \tau_n^{-2}\sum_{i=1}^{N(n)} \Big(E^*(A_{n,i}^2)^{1+\eta/2}\Big)^{1/(1+\eta/2)} \cdot (P^*\{|A_{n,i}/\tau_n|^{1+\eta/2} > \varepsilon^{1+\eta/2}\})^{(\eta/2)/(1+\eta/2)}\\
&\le \tau_n^{-(2+\eta)}\varepsilon^{-\eta}\sum_{i=1}^{N(n)} E^*(|A_{n,i}|^{2+\eta})\\
&= \tau_n^{-(2+\eta)}\varepsilon^{-\eta}\,N(n)\,a(n)^{1+\eta/2}\cdot\mathcal{O}_P(1),
\end{aligned} \tag{3.53}$$

where (3.52) was used in the final step. From (3.41) and (3.51) we obtain

$$(\tau_n^2/n\delta)^{-(1+\eta/2)} = \Big(v^{*2}\cdot\frac{\tau_n^2}{v^{*2}\,n\delta}\Big)^{-(1+\eta/2)} \xrightarrow{P} (\underline{c}^T\Sigma_M\,\underline{c})^{-(2+\eta)},$$

which yields

$$(\tau_n^2/n\delta)^{-(1+\eta/2)} = \mathcal{O}_P(1). \tag{3.54}$$

From the definitions of $N(n), a(n)$ and $b(n)$ we have $N(n) = 2n\delta/(a(n)+b(n))$ and (3.44) ensures that $a(n)/(a(n)+b(n)) = \mathcal{O}(1)$. Therefore, the right-hand side of (3.53) equals

$$\begin{aligned}
&\tau_n^{-(2+\eta)}\varepsilon^{-\eta}\,\frac{2n\delta}{a(n)}\cdot\frac{a(n)}{a(n)+b(n)}\,a(n)^{1+\eta/2}\cdot\mathcal{O}_P(1)\\
&= \tau_n^{-(2+\eta)}\varepsilon^{-\eta}\,(n\delta)\,a(n)^{\eta/2}\cdot\mathcal{O}_P(1)
\end{aligned}$$

$$
\begin{aligned}
&= (\tau_n^2/n\delta)^{-(1+\eta/2)} \left(a(n)/n\delta\right)^{\eta/2} \cdot \mathcal{O}_P(1) \\
&= (\tau_n^2/n\delta)^{-(1+\eta/2)} \left(\frac{2}{N(n)} \frac{a(n)}{a(n)+b(n)}\right)^{\eta/2} \cdot \mathcal{O}_P(1) \\
&= \mathcal{O}_P(1) \cdot o(1) \cdot \mathcal{O}_P(1) \\
&= o_P(1)
\end{aligned}
$$

since $N(n) \to \infty$ as $n \to \infty$, and because of (3.54). Hence, the Lindeberg condition (3.49) is fulfilled and (3.50) holds. Together with (3.51) this yields (3.48). The decomposition (3.46) combined with (3.47) and (3.48) now gives (3.42) which, together with Slutsky's Theorem, yields

$$
\underline{c}^T \underline{L}_{n,M}^*(x) \xrightarrow{d^*} \mathcal{N}(0, \underline{c}^T \Sigma_M \, \underline{c}) \quad \text{in } P\text{-prob. } \forall \, \underline{c} \in \mathbb{R}^q. \tag{3.55}
$$

We finish the proof by showing that first truncating at M and then letting $M \to \infty$ does not alter the asymptotics. Keeping in mind that $\underline{c}^T \Sigma_M \, \underline{c} \to \underline{c}^T \Sigma \, \underline{c}$ as $M \to \infty$, cf. (3.35), we can apply Corollary 7.7.1 from Anderson (1971) to (3.55). Therefore, if we can show that

$$
E^*\Big((\underline{c}^T \underline{L}_n^*(x) - \underline{c}^T \underline{L}_{n,M}^*(x))^2 \Big) = \mathcal{O}_P(1/M), \ \text{ uniformly in } n, \tag{3.56}
$$

holds, we get

$$
\underline{c}^T \underline{L}_n^*(x) \xrightarrow{d^*} \mathcal{N}(0, \underline{c}^T \Sigma \, \underline{c}) \quad \text{in } P\text{-prob. } \forall \, \underline{c} \in \mathbb{R}^q. \tag{3.57}
$$

from Anderson's result. In order to prove (3.56) we first establish two auxiliary assertions. On the one hand, since K is a bounded function, there exists a constant C such that

$$
\sum_{t=1}^{n-|h|} (n\delta)^{-1} K\Big(\frac{x - t/n}{\delta}\Big) \, K\Big(\frac{x - (t+|h|)/n}{\delta}\Big) \le C \qquad \forall \, |h| \le n, \, \forall \, n \in \mathbb{N},
$$

because no more than $2n\delta$ summands in this expression are non-zero. On the other hand the strict stationarity of $(\underline{X}_t^*)$ and $(\underline{X}_{t,M}^*)$ guarantees that $(\underline{V}_t^*)_{t \in \mathbb{Z}}$ with $\underline{V}_t^* := \underline{X}_t^* - \underline{X}_{t,M}^* = \sum_{j=M+1}^{\infty} \widehat{B}_j(p) \, \underline{\varepsilon}_{t-j}^*$ fulfils

$$
\mathrm{Cov}^*\Big(\underline{c}^T \underline{V}_{t+|h|}^*, \underline{c}^T \underline{V}_t^*\Big) = \mathrm{Cov}^*\Big(\underline{c}^T \underline{V}_{|h|}^*, \underline{c}^T \underline{V}_0^*\Big) \quad \forall \, t \in \mathbb{Z}.
$$

Using these assertions and again the abbreviation $K_{t,x} := K((x - t/n)/\delta)$ we can derive in a straightforward way

$$
M \cdot E^*\Big((\underline{c}^T \underline{L}_n^*(x) - \underline{c}^T \underline{L}_{n,M}^*(x))^2 \Big)
$$

$$
\begin{aligned}
&= M \cdot E^*\Big(\Big(\underline{c}^T (n\delta)^{-1/2} \sum_{t=1}^{n} K_{t,x} \cdot (\underline{X}_t^* - \underline{X}_{t,M}^*)\Big)^2\Big) \\
&= M \cdot \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} (n\delta)^{-1} K_{t,x}\, K_{t+|h|,x}\, E^*\Big(\underline{c}^T \underline{V}_{t+|h|}^* \cdot \underline{c}^T \underline{V}_t^*\Big) \\
&\leq C \cdot M \cdot \sum_{h=-(n-1)}^{n-1} \operatorname{Cov}^*\Big(\underline{c}^T \underline{V}_{|h|}^* \cdot \underline{c}^T \underline{V}_0^*\Big) \\
&\leq 2C \cdot M \cdot \sum_{h=0}^{n-1} \Big|\operatorname{Cov}^*\Big(\underline{c}^T \underline{V}_h^*, \underline{c}^T \underline{V}_0^*\Big)\Big| \qquad (3.58)
\end{aligned}
$$

Decomposing

$$
\underline{c}^T \underline{V}_h^* = \sum_{j=M+1}^{M+h} \underline{c}^T\, \widehat{B}_j(p)\, \underline{\varepsilon}_{h-j}^* + \sum_{j=M+h+1}^{\infty} \underline{c}^T\, \widehat{B}_j(p)\, \underline{\varepsilon}_{h-j}^*,
$$

and noting that the first summand on the right-hand side is (conditionally) independent of $\underline{c}^T \underline{V}_0^*$, the right-hand side of (3.58) can be bounded by

$$
\begin{aligned}
&2C \cdot M \cdot \sum_{h=0}^{n-1} \Big\| \sum_{j=M+h+1}^{\infty} \underline{c}^T\, \widehat{B}_j(p)\, \underline{\varepsilon}_{h-j}^* \Big\|_{*2} \cdot \Big\| \underline{c}^T \underline{V}_0^* \Big\|_{*2} \\
&\leq 2C \cdot M \cdot \sum_{h=0}^{n-1} \sum_{r,s=1}^{q} |c_r| \sum_{j=M+h+1}^{\infty} \Big|\widehat{B}_j(p)^{(r,s)}\Big| \cdot \Big\| \varepsilon_{h-j}^*(s) \Big\|_{*2} \cdot \Big\| \underline{c}^T \underline{V}_0^* \Big\|_{*2} \\
&\leq \mathcal{O}_P(1) \cdot M \cdot \sum_{h=0}^{n-1} \sum_{j=M+h+1}^{\infty} \Big\|\widehat{B}_j(p)\Big\| \cdot \Big\| \underline{c}^T \underline{V}_0^* \Big\|_{*2},
\end{aligned}
$$

where (3.3) was used as well as the fact that $\|\varepsilon_t^*(s)\|_{*2} = \mathcal{O}_P(1)$, uniformly in t and s, follows from (3.27). Proceeding in the same way for $\|\underline{c}^T \underline{V}_0^*\|_{*2}$ the right-hand side can be bounded by

$$
\begin{aligned}
&\mathcal{O}_P(1) \cdot M \cdot \sum_{h=0}^{n-1} \sum_{j=M+h+1}^{\infty} \Big\|\widehat{B}_j(p)\Big\| \cdot \mathcal{O}_P(1) \cdot \sum_{k=M+1}^{\infty} \Big\|\widehat{B}_k(p)\Big\| \\
&\leq \mathcal{O}_P(1) \cdot \sum_{h=0}^{\infty} \sum_{j=M+h+1}^{\infty} \Big\|\widehat{B}_j(p)\Big\| \cdot \sum_{k=M+1}^{\infty} k \cdot \Big\|\widehat{B}_k(p)\Big\| \\
&\leq \mathcal{O}_P(1) \cdot \sum_{j=M+1}^{\infty} (j-M) \cdot \Big\|\widehat{B}_j(p)\Big\| \cdot \sum_{k=1}^{\infty} k \cdot \Big\|\widehat{B}_k(p)\Big\| \\
&\leq \mathcal{O}_P(1) \cdot \Big(\sum_{j=1}^{\infty} j \cdot \Big\|\widehat{B}_j(p)\Big\|\Big)^2 \\
&\leq \mathcal{O}_P(1),
\end{aligned}
$$

using (3.26). This yields (3.56) and therefore (3.57). Applying the Cramér-Wold device, we get

$$\underline{L}_n^*(x) \xrightarrow{d^*} \mathcal{N}(\underline{0}, \Sigma) \quad \text{in } P\text{-probability} , \tag{3.59}$$

which completes the proof. □

3.8 Proofs of the auxiliary results

Proof of Lemma 3.1:

For a process fulfilling Assumption 5 the assumptions of Theorem 6.6.12 in Hannan and Deistler (1988) are met. In particular, we derived in (3.12) the invertibilty of the power series of the optimal factor on the entire unit disk. Thus, the Theorem applies, but it is worth mentioning that Hannan and Deistler use a different weight function. However, inspecting the proof shows that it is valid for any weight function that fulfils $\nu(j) \leq \nu(k)\,\nu(j-k)$ for all $j, k \in \mathbb{Z}$ and our weight function has this property as is shown in (3.6). Also, it should be noted that the original version of the Theorem contains additional weighting factors Σ^{-1} and Σ_p^{-1} but according to Remark 3, p. 270 in Hannan and Deistler (1988), the inequality is also true without these factors. □

Proof of Lemma 3.2:

First, we show that $\det A_p(z) \neq 0 \quad \forall\, |z| \leq 1$. This was proven by Whittle (1963), see chapter 3 there, under less restrictive assumptions than we apply in this Lemma, except that Whittle's proof needs the matrix $G(p)$ to be (strictly) positive definite. We will show that this condition follows from our assumptions.

According to Brockwell and Davis (1991), p. 393, $G(p)$ is always positive semi-definite and hemitian. Therefore, all of its eigenvalues are non-negative. Since $\det G(p)$ is the product of its eigenvalues and $G(p)$ is always invertible under the imposed conditions, cf. section 3.3, it follows immediately that all eigenvalues are strictly positive and $G(p)$ is positive definite. Hence, we can apply Whittle's proof and get

$$\det A_p(z) \neq 0 \quad \forall\, |z| \leq 1. \tag{3.60}$$

With the rest of this proof we will strengthen this assertion as in Lemma 3.2. We start with two short auxiliary results. Firstly, it holds uniformly for all $|z| \leq 1$

$$\|A_p(z) - A(z)\| \;=\; \Big\| \sum_{k=1}^{p} (A_k - A_k(p)) z^k + \sum_{k=p+1}^{\infty} A_k z^k \Big\|$$

$$\leq \sum_{k=1}^{p} \|A_k - A_k(p)\| + \sum_{k=p+1}^{\infty} \|A_k\|$$
$$\leq \sum_{k=1}^{p} (1+|k|)^r \|A_k - A_k(p)\| + \sum_{k=p+1}^{\infty} \|A_k\|$$

and the right-hand side converges to zero as $p \to \infty$ because of Lemma 3.1. Thus, we get

$$\|A_p(z) - A(z)\| \longrightarrow 0 \quad \text{uniformly } \forall\,|z| \leq 1 \text{ as } p \to \infty. \tag{3.61}$$

Secondly, the matrix norm $\|\cdot\|$ fulfils (3.3) and therefore convergence of matrices in the sense of the matrix norm is equivalent to convergence of all entries. Since all entries of $A(z)$ are continuous functions in z, it follows for any sequence (z_k) in the closed unit disk with $z_k \to z_0$ as $k \to \infty$ that

$$\|A(z_0) - A(z_k)\| \longrightarrow 0 \qquad \text{as } k \to \infty. \tag{3.62}$$

Now assume that Lemma 3.2 was not true. Then there is a sequence $p(k)$ with $p(k) \to \infty$ as $k \to \infty$ and a sequence of complex numbers (z_k) with $|z_k| \leq 1 + \frac{1}{p(k)}$ such that

$$\det\, A_{p(k)}(z_k) \longrightarrow 0, \quad k \to \infty. \tag{3.63}$$

Furthermore, assume there is a subsequence of (z_k) completely inside the closed unit disk. Without loss of generality let this subsequence be (z_k) itself. As it holds $\det A_{p(k)}(z) \neq 0$ for all $|z| \leq 1$, the function $f(z) = 1/(\det A_{p(k)}(z))$ is holomorphic on the open unit disk and we can apply the maximum principle for holomorphic functions: f takes its maximum absolute value on the boundary of the disk, i.e.

$$\frac{1}{|\det A_{p(k)}(z)|} \leq \max_{|\tilde{z}|=1} \frac{1}{|\det A_{p(k)}(\tilde{z})|} \quad \forall\,|z| \leq 1,$$

which is equivalent to

$$|\det A_{p(k)}(z)| \geq \min_{|\tilde{z}|=1} |\det A_{p(k)}(\tilde{z})| \quad \forall\,|z| \leq 1.$$

As the function $|\det A_{p(k)}(\cdot)|$ is continuous, there is a $z^*_{p(k)}$ with $|z^*_{p(k)}| = 1$ such that

$$|\det A_{p(k)}(z^*_{p(k)})| = \min_{|\tilde{z}|=1} |\det A_{p(k)}(\tilde{z})| = \min_{|\tilde{z}|\leq 1} |\det A_{p(k)}(\tilde{z})|. \tag{3.64}$$

Since $(z^*_{p(k)})$ is a sequence on the unit circle, it possesses a convergent subsequence. Again, without loss of generality, let $(z^*_{p(k)})$ itself be this subsequence and denote its limit by $z^* := \lim_{k\to\infty} z^*_{p(k)}$. Combining (3.63) and (3.64) we get

$$|\det A_{p(k)}(z^*_{p(k)})| \leq |\det A_{p(k)}(z_k)| \longrightarrow 0, \quad \text{as } k \to \infty. \tag{3.65}$$

As $z^*_{p(k)} \to z^*$, (3.62) yields $\|A(z^*_{p(k)}) - A(z^*)\| \to 0$. Also, as $|z^*_{p(k)}| = 1$, (3.61) yields

$$\|A_{p(k)}(z^*_{p(k)}) - A(z^*_{p(k)})\| \to 0.$$

With these two assertions we get

$$\begin{aligned}\|A_{p(k)}(z^*_{p(k)}) - A(z^*)\| &= \|A_{p(k)}(z^*_{p(k)}) - A(z^*_{p(k)}) + A(z^*_{p(k)}) - A(z^*)\| \\ &\leq \|A_{p(k)}(z^*_{p(k)}) - A(z^*_{p(k)})\| + \|A(z^*_{p(k)}) - A(z^*)\| \to 0.\end{aligned}$$

As convergence in matrix norm implies convergence of all entries, and since it is well known that this yields convergence of the corresponding determinants, it follows

$$\det A(z^*) = \lim_{k\to\infty} \det A_{p(k)}(z^*_{p(k)})$$

Applying (3.65) to this equation, we immediately get $\det A(z^*) = 0$, which contradicts (3.13) because $|z^*| = 1$. Therefore, the assumption is not true and there is no such subsequence (z_k) completely inside the closed unit disk.
As we still assume Lemma 3.2 was not true there has to be a sequence (z_k) fulfilling (3.63). We have proven that there cannot be a subsequence of (z_k) staying completely inside the closed unit disk. Therefore, there has to be a subsequence lying completely in the ring $1 < |z_k| \leq 1 + \frac{1}{p(k)}$. Such a sequence has a convergent subsequence and, again without loss of generality, let this be (z_k) itself, with $z_k \to z_0$ as $k \to \infty$. Then it holds obviously $|z_0| = 1$. We will show that this also leads to a contradiction.
We have

$$\begin{aligned}\Big\| \sum_{j=1}^{p(k)} (A_j(p(k)) - A_j)\, z_k^j \Big\| &\leq |z_k|^{p(k)} \sum_{j=1}^{p(k)} \|A_j(p(k)) - A_j\| \\ &\leq \Big|1 + \frac{1}{p(k)}\Big|^{p(k)} \cdot \sum_{j=1}^{p(k)} \|A_j(p(k)) - A_j\|,\end{aligned}$$

where the left factor is bounded in k and the right factor converges to zero as $k \to \infty$ by Lemma 3.1. Hence, we get

$$\Big\| \sum_{j=1}^{p(k)} (A_j(p(k)) - A_j)\, z_k^j \Big\| \longrightarrow 0, \quad \text{as } k \to \infty. \tag{3.66}$$

A similar result holds for the expression $\sum_{j=1}^{p(k)} A_j(z_k^j - z_0^j)$. Denote an arbitrary entry of A_j by $A_j^{(r,s)}$. Then it holds $\sum_{j=1}^{\infty} |A_j^{(r,s)}| < \infty$ because of (3.3). Thus, applying the dominated convergence theorem leads to

$$\sum_{j=1}^{p(k)} A_j^{(r,s)}(z_k^j - z_0^j) \longrightarrow 0, \quad \text{as } k \to \infty,$$

which, again using (3.3), yields

$$\Big\| \sum_{j=1}^{p(k)} A_j(z_k^j - z_0^j) \Big\| \longrightarrow 0, \quad \text{as } k \to \infty. \tag{3.67}$$

We can now derive

$$\begin{aligned}
& \|A_{p(k)}(z_k) - A(z_0)\| \\
= \;& \Big\| I - \sum_{j=1}^{p(k)} A_j(p(k))\, z_k^j - I + \sum_{j=1}^{\infty} A_j\, z_0^j \Big\| \\
= \;& \Big\| \sum_{j=1}^{p(k)} [A_j z_0^j - A_j(p(k)) z_k^j] + \sum_{j=p(k)+1}^{\infty} A_j\, z_0^j \Big\| \\
\leq \;& \Big\| \sum_{j=1}^{p(k)} [A_j - A_j(p(k))]\, z_k^j \Big\| + \Big\| \sum_{j=1}^{p(k)} A_j\, (z_0^j - z_k^j) \Big\| + \sum_{j=p(k)+1}^{\infty} \|A_j\|,
\end{aligned}$$

where the right-hand side tends to zero as $k \to \infty$, according to (3.66), (3.67) and (3.9). Therefore, the corresponding determinants converge and we get with (3.63)

$$\det A(z_0) = \lim_{k\to\infty} \det A_{p(k)}(z_k) = 0$$

This is again a contradiction to (3.13). Hence, there is no sequence fulfilling (3.63) and the assertion of Lemma 3.2 holds true. □

Proof of Lemma 3.3:

We first introduce a subspace of the space of all sequences of $q \times q$-matrices: For some fixed $r \geq 0$ and a fixed matrix norm $\|\cdot\|$, define

$$l_q := \{(A_n)_{n\in\mathbb{N}_0} \Big| A_n \in \mathbb{R}^{q\times q} \text{ and } \sum_{j=0}^{\infty} (1+|j|)^r \|A_j\| < \infty\}.$$

In this vector space, addition of two elements is defined as $(A_n) + (B_n) = (C_n)$ with $C_n := A_n + B_n$ for all $n \in \mathbb{N}_0$ (where addition of two matrices is the usual entrywise addition). Likewise, multiplication with scalars is the multiplication of each member of the sequence with the scalar (and entrywise multiplication for each member). The natural norm of this space is

$$\| \, (A_n) \, \|_{l_q} := \sum_{j=0}^{\infty} (1+|j|)^r \|A_j\|.$$

We define multiplication of two vectors as convolution of the two sequences:

$$(A_n) \cdot (B_n) = (C_n), \quad \text{where} \;\; C_n := \sum_{k=0}^{n} A_k B_{n-k} \;\; \forall \, n \in \mathbb{N}_0.$$

One can easily verify that this multiplication of vectors together with the norm $\| \cdot \|_{l_q}$ is submultiplicative, i.e.

$$\| \, (A_n) \cdot (B_n) \, \|_{l_q} \leq \| \, (A_n) \, \|_{l_q} \cdot \| \, (B_n) \, \|_{l_q}.$$

Thus, the space l_q is indeed closed under the defined multiplication. It is also obvious that $(\mathcal{I}_n) := (I, 0, 0, \ldots)$ is the identity element of the multiplication as it holds for any vector $(A_n) \in l_q$ that

$$(A_n) \cdot (\mathcal{I}_n) = (\mathcal{I}_n) \cdot (A_n) = (A_n).$$

We now turn our attention to some particular elements of l_q which appear in Lemma 3.3. Let $A_1, A_2, \ldots$ be the autoregressive and $B_1, B_2, \ldots$ be the moving average coefficients of our stochastic process as defined in (3.5). We define the following sequences of matrices

$$(\mathcal{A}_n) := (I, -A_1, -A_2, \ldots) \;\; \text{and} \;\; (\mathcal{B}_n) := (I, B_1, B_2, \ldots)$$

which are both elements of l_q because of (3.9). Now recall from the definition of the coefficients A_k and B_k from section 3.2 that $(\mathcal{A}_n)$ and $(\mathcal{B}_n)$ can actually be expressed as

$$(\mathcal{A}_n) = (\Sigma^{1/2} D_0, \Sigma^{1/2} D_1, \Sigma^{1/2} D_2, \ldots) \;\; \text{and} \;\; (\mathcal{B}_n) = (C_0 \Sigma^{-1/2}, C_1 \Sigma^{-1/2}, C_2 \Sigma^{-1/2}, \ldots)$$

where the C_k and D_k are the Fourier coefficients of ϕ and ϕ^{-1}, respectively, cf. (3.8). Therefore, we can derive

$$\begin{aligned} (\mathcal{B}_n) \cdot (\mathcal{A}_n) &= \left(\sum_{k=0}^{n} C_k \Sigma^{-1/2} \Sigma^{1/2} D_{n-k} \right)_{n \in \mathbb{N}_0} \\ &= \left(\sum_{k=0}^{n} C_k D_{n-k} \right)_{n \in \mathbb{N}_0} = (I, 0, 0, \ldots) = (\mathcal{I}_n), \end{aligned}$$

where the next-to-last equation follows directly from (3.11). Although matrix multiplication, as well as the multiplication in l_q, is not commutative, the same result follows for $(\mathcal{A}_n) \cdot (\mathcal{B}_n)$ because $\phi\phi^{-1} = \phi^{-1}\phi = I$. Hence, $(\mathcal{A}_n)$ and $(\mathcal{B}_n)$ are inverse in l_q and we denote $(\mathcal{B}_n) = (\mathcal{A}_n)^{-1}$.

We now focus on the finite predictor coefficients which appear as the coefficients in $A_p(z) = I - \sum_{k=1}^{p} A_k(p) z^k$, cf. section 3.3. Since $\det A_p(z)$ is continuous in z and uniformly bounded away from zero on the entire region $|z| \leq 1 + (1/p)$ according to Lemma 3.2 (recall that here, we are only looking at $p \geq p_1$), it holds

$$\det A_p(z) \neq 0 \qquad \forall\ |z| \leq 1 + (1/p) + \varepsilon$$

for some $\varepsilon > 0$. As each entry of $A_p(z)$ is a polynomial, each entry of $A_p(z)^{-1}$ possesses an absolutely convergent power series expansion in the interior of the region $|z| \leq 1 + (1/p) + \varepsilon$, i.e. particularly on $|z| \leq 1 + (1/p)$. Denote this expansion, for an arbitrary entry (s,t) of $A_p(z)^{-1}$, by $\sum_{k=0}^{\infty} b_k^{(s,t)}(p)\, z^k$. From the absolute summability on $|z| \leq 1 + (1/p)$ we immediately get the exponential decay $|b_k^{(s,t)}(p)| \leq \mathcal{O}([1 + (1/p)]^{-k})$. The matrices $B_k(p)$ from section 3.3 are obviously given by $B_k(p) = (b_k^{(s,t)}(p))_{1 \leq s,t \leq q}$ and we can define the matrix sequences

$$\begin{aligned} (\mathcal{A}_n^p) &:= (I, -A_1(p), -A_2(p), \ldots, -A_p(p), 0, 0, \ldots) \\ (\mathcal{B}_n^p) &:= (B_0(p), B_1(p), \ldots). \end{aligned}$$

Clearly, $(\mathcal{A}_n^p)$ is an element of l_q as it consists of only a finite number of non-zero-matrices. We also get

$$\| (\mathcal{B}_n^p) \|_{l_q} \leq \sum_{s,t=1}^{q} \sum_{k=0}^{\infty} (1 + |k|)^r\, |b_k^{(s,t)}(p)| < \infty$$

from the exponential decay of the $b_k^{(s,t)}(p)$. The definition of $(\mathcal{B}_n^p)$ yields $(\mathcal{B}_n^p) \cdot (\mathcal{A}_n^p) = (\mathcal{A}_n^p) \cdot (\mathcal{B}_n^p) = (\mathcal{I}_n)$. Therefore, we can conclude that

$$(\mathcal{A}_n), (\mathcal{B}_n), (\mathcal{A}_n^p), (\mathcal{B}_n^p) \in l_q, \quad (\mathcal{B}_n) = (\mathcal{A}_n)^{-1}, \quad (\mathcal{B}_n^p) = (\mathcal{A}_n^p)^{-1}. \tag{3.68}$$

We now have established the necessary preliminary results to prove the assertion of Lemma 3.3. Using (3.68) and the submultiplicativity of $\| \cdot \|_{l_q}$ we derive

$$\begin{aligned} & \sum_{k=0}^{\infty} (1 + |k|)^r\, \|B_k(p) - B_k\| \\ = \quad & \| (\mathcal{B}_n^p) - (\mathcal{B}_n) \|_{l_q} \qquad (3.69) \\ = \quad & \| (\mathcal{B}_n^p) \cdot [(\mathcal{A}_n) - (\mathcal{A}_n^p)] \cdot (\mathcal{B}_n) \|_{l_q} \end{aligned}$$

$$\begin{aligned}
&\le \ \| (\mathcal{B}_n^p) \|_{l_q} \cdot \| (\mathcal{A}_n) - (\mathcal{A}_n^p) \|_{l_q} \cdot \| (\mathcal{B}_n) \|_{l_q} \\
&\le \ \Big(\| (\mathcal{B}_n^p) - (\mathcal{B}_n) \|_{l_q} + \| (\mathcal{B}_n) \|_{l_q} \Big) \cdot \| (\mathcal{A}_n) - (\mathcal{A}_n^p) \|_{l_q} \cdot \| (\mathcal{B}_n) \|_{l_q}. \qquad (3.70)
\end{aligned}$$

From Baxter's inequality, cf. Lemma 3.1, we can infer

$$\begin{aligned}
\| (\mathcal{A}_n) - (\mathcal{A}_n^p) \|_{l_q} &= \sum_{k=1}^{p} (1+|k|)^r \, \|A_k(p) - A_k\| + \sum_{k=p+1}^{\infty} (1+|k|)^r \, \|A_k\| \\
&\le C \cdot \sum_{k=p+1}^{\infty} (1+|k|)^r \, \|A_k\| \qquad \forall\, p \ge p_1, \qquad (3.71)
\end{aligned}$$

since $p_1 \ge p_0$. Hence, one can always find $p \in \mathbb{N}$ such that $\| (\mathcal{A}_n) - (\mathcal{A}_n^p) \|_{l_q}$ becomes arbitrarily small. In particular, for some arbitrary $\delta \in (0,1)$, choose $p_2 \ge p_1$ such that

$$\| (\mathcal{A}_n) - (\mathcal{A}_n^p) \|_{l_q} \le \delta / \| (\mathcal{B}_n) \|_{l_q}$$

for all $p \ge p_2$. Then we can subtract (3.70) from (3.69) and get

$$\begin{aligned}
\| (\mathcal{B}_n^p) - (\mathcal{B}_n) \|_{l_q} &\le \frac{\| (\mathcal{B}_n) \|_{l_q}^2 \cdot \| (\mathcal{A}_n) - (\mathcal{A}_n^p) \|_{l_q}}{1 - \| (\mathcal{A}_n) - (\mathcal{A}_n^p) \|_{l_q} \cdot \| (\mathcal{B}_n) \|_{l_q}} \\
&\le \frac{\| (\mathcal{B}_n) \|_{l_q}^2}{1-\delta} \cdot \| (\mathcal{A}_n) - (\mathcal{A}_n^p) \|_{l_q} \qquad (3.72)
\end{aligned}$$

for all $p \ge p_2$. Since the first factor on the right-hand side of (3.72) does not depend on p and is bounded, applying (3.71) to the second factor yields

$$\sum_{k=1}^{\infty} (1+|k|)^r \, \|B_k(p) - B_k\| \le C \cdot \sum_{k=p+1}^{\infty} (1+|k|)^r \, \|A_k\| \qquad \forall\, p \ge p_2,$$

which completes the proof. □

Proof of Lemma 3.4:

Denote the rank of a matrix A by $\mathrm{rk}(A)$. It is easy to see from the definitions of $\widehat{G}(p)$ and $\widehat{\Gamma}(h)$ that $\widehat{G}(p) = \frac{1}{n} H H^T$. Since $\widehat{G}(p)$ is a $pq \times pq$-matrix it follows

$$\begin{aligned}
\det \widehat{G}(p) \ne 0 \ &\Leftrightarrow \ \mathrm{rk}(\widehat{G}(p)) = pq \\
&\Leftrightarrow \ \mathrm{rk}(H H^T) = pq \\
&\Leftrightarrow \ \mathrm{rk}(H) = pq.
\end{aligned}$$

H is a $qp \times (n+p-1)$-matrix. If $qp > n+p-1$, it follows $\mathrm{rk}(H) \le n+p-1 < pq$ and $\widehat{G}(p)$ is singular. Now assume $qp \le n+p-1$. Then it is $\mathrm{rk}(H) = pq$ if and only if the rows of H are linearly independent which completes the proof. □

Proof of Lemma 3.8:
Note that Assumption 6 guarantees

$$\begin{aligned}\sup_{|z|\leq 1+(1/p)} \|\widehat{A}_p(z) - A_p(z)\| &\leq \sup_{|z|\leq 1+(1/p)} \sum_{k=1}^{p} \|\widehat{A}_k(p) - A_k(p)\| \cdot |z|^k \\ &\leq (1+(1/p))^p \cdot \sum_{k=1}^{p} \|\widehat{A}_k(p) - A_k(p)\| \\ &\leq 3 \cdot \frac{1}{p^2} \cdot \mathcal{O}_P(1).\end{aligned}$$

In the following we use the obvious notation $C^{(r,s)}$ for the (r,s)-entry of any matrix C. From (3.3) we can infer the same rate of decrease for each entry of the respective matrix:

$$\sup_{|z|\leq 1+(1/p)} |\widehat{A}_p(z)^{(r,s)} - A_p(z)^{(r,s)}| = \frac{1}{p^2} \cdot \mathcal{O}_P(1). \tag{3.73}$$

Since determinants are continuous functions of the entries of the respective matrices, we also immediately get

$$\sup_{|z|\leq 1+(1/p)} |\det \widehat{A}_p(z) - \det A_p(z)| = \frac{1}{p^2} \cdot \mathcal{O}_P(1). \tag{3.74}$$

It was shown in Lemma 3.2 that, for $p \geq p_1$, $\det A_p(z)$ is uniformly bounded away from zero on the region $|z| \leq 1 + (1/p)$. Because of (3.74), for p large enough, $\det \widehat{A}_p(z)$ is also bounded away from zero in probability on the very same region. In particular, we can choose $p_3 \in \mathbb{N}$ such that

$$|\det \widehat{A}_p(z)| \geq \delta_0 \text{ in prob.}, \ |\det A_p(z)| \geq \delta_0 \quad \forall\, p \geq p_3, \forall\, |z| \leq 1 + (1/p) \tag{3.75}$$

for some $\delta_0 > 0$. Therefore, the power series expansion in (3.21) is actually valid (in probability) for all $|z| \leq 1 + (1/p)$. Since the k-th coefficient of the power series expansion of entry $(\widehat{A}_p(z)^{-1} - A_p(z)^{-1})^{(r,s)}$ is given by $\widehat{B}_k(p)^{(r,s)} - B_k(p)^{(r,s)}$, we use (3.3) and Cauchy's inequality for holomorphic functions to derive

$$\begin{aligned}\|\widehat{B}_k(p) - B_k(p)\| &\leq \sum_{r,s=1}^{q} |\widehat{B}_k(p)^{(r,s)} - B_k(p)^{(r,s)}| \\ &\leq \sum_{r,s=1}^{q} \left(1+\frac{1}{p}\right)^{-k} \max_{|z|=1+(1/p)} |(\widehat{A}_p(z)^{-1} - A_p(z)^{-1})^{(r,s)}| \\ &\leq q^2 \left(1+\frac{1}{p}\right)^{-k} \max_{|z|=1+(1/p)} \|\widehat{A}_p(z)^{-1} - A_p(z)^{-1}\| \quad \text{in prob.}\end{aligned}$$

Thus, the proof can be completed by showing

$$\max_{|z|=1+(1/p)} \|\widehat{A}_p(z)^{-1} - A_p(z)^{-1}\| = \frac{1}{p^2} \cdot \mathcal{O}_P(1). \tag{3.76}$$

We will do this by using adjoint matrices. Recall that for any regular matrix C it holds $C^{-1} = (\det C)^{-1}\, C_{adj}$ where C_{adj} is the adjoint matrix of C, i.e. the $q\times q$-matrix with entries $C_{adj}^{(r,s)} = (-1)^{r+s} \det C^{(-s,-r)}$. Here, $C^{(-s,-r)}$ is the $(q-1)\times(q-1)$-matrix which results from removing row s and column r from matrix C. From this definition we get

$$\begin{aligned}
& \|\widehat{A}_p(z)^{-1} - A_p(z)^{-1}\| \\
= \;& \Big\| \frac{1}{\det \widehat{A}_p(z)} \widehat{A}_p(z)_{adj} - \frac{1}{\det A_p(z)} A_p(z)_{adj} \Big\| \\
\leq \;& \frac{1}{|\det \widehat{A}_p(z)|} \Big\| \widehat{A}_p(z)_{adj} - A_p(z)_{adj} \Big\| + \Big| \frac{1}{\det \widehat{A}_p(z)} - \frac{1}{\det A_p(z)} \Big| \cdot \Big\| A_p(z)_{adj} \Big\| \\
=: \;& I(p,z) + II(p,z).
\end{aligned}$$

Therefore,

$$\max_{|z|=1+(1/p)} \|\widehat{A}_p(z)^{-1} - A_p(z)^{-1}\| \leq \max_{|z|=1+(1/p)} I(p,z) + \max_{|z|=1+(1/p)} II(p,z). \tag{3.77}$$

In the remainder of this proof we bound the two summands on the right-hand side of (3.77) in probability, starting with the first one.
For all $p \geq p_3$ and any $|z| = 1 + (1/p)$ we can use (3.75) to get

$$\begin{aligned}
I(p,z) \;&\leq\; \frac{1}{\delta_0} \|\widehat{A}_p(z)_{adj} - A_p(z)_{adj}\| \\
&\leq\; \frac{1}{\delta_0} \sum_{r,s=1}^{q} |\widehat{A}_p(z)_{adj}^{(r,s)} - A_p(z)_{adj}^{(r,s)}| \\
&=\; \frac{1}{\delta_0} \sum_{r,s=1}^{q} |(-1)^{r+s} \det \widehat{A}_p(z)^{(-s,-r)} - (-1)^{r+s} \det A_p(z)^{(-s,-r)}| \\
&\leq\; \frac{1}{\delta_0} \sum_{r,s=1}^{q} \sup_{|z|\leq 1+(1/p)} |\det \widehat{A}_p(z)^{(-s,-r)} - \det A_p(z)^{(-s,-r)}| \quad \text{ in prob.},
\end{aligned}$$

where all q^2 summands on the right-hand side can be bounded by $p^{-2}\mathcal{O}_P(1)$ with the very same arguments as in (3.74). Since this bound is independent of z, we get

$$\max_{|z|=1+(1/p)} I(p,z) \leq \frac{1}{p^2} \cdot \mathcal{O}_P(1). \tag{3.78}$$

As for the second summand of (3.77), we first establish a bound for $\|A_p(z)_{adj}\|$. For all $p \geq p_3$ and any $|z| = 1+(1/p)$ we can use (3.3) and Baxter's inequality for $r = 0$, cf. Lemma 3.1, to bound each entry of $A_p(z)$ by

$$\begin{aligned}
\|A_p(z)\| &\leq \sum_{k=1}^{p} \|A_k(p)\| \cdot |z|^k \\
&\leq (1+(1/p))^p \cdot \left(\sum_{k=1}^{p} \|A_k\| + \sum_{k=1}^{p} \|A_k(p) - A_k\| \right) \\
&\leq 3 \cdot \left(\sum_{k=1}^{\infty} \|A_k\| + C \cdot \sum_{k=p+1}^{\infty} \|A_k\| \right) \\
&\leq C' \cdot \sum_{k=1}^{\infty} \|A_k\| < \infty.
\end{aligned}$$

Note that this bound is uniform for all $p \geq p_3$ and all $|z| = 1+(1/p)$. Therefore, as the determinant is a continuous function of the entries, there exists $C'' < \infty$ such that it holds for all $1 \leq r, s \leq q$:

$$|\det A_p(z)^{(-s,-r)}| \leq C'' \qquad \forall\, p \geq p_3, \quad \forall\, |z| = 1+(1/p).$$

This yields the desired bound via

$$\begin{aligned}
\|A_p(z)_{adj}\| &\leq \sum_{r,s=1}^{q} |A_p(z)_{adj}^{(r,s)}| \\
&= \sum_{r,s=1}^{q} |\det A_p(z)^{(-s,-r)}| \\
&\leq q^2 \cdot C'' \qquad\qquad \forall\, p \geq p_3, \quad \forall\, |z| = 1+(1/p). \qquad (3.79)
\end{aligned}$$

We now turn back to the second summand of (3.77). For all $p \geq p_3$ and any $|z| = 1+(1/p)$ we can use (3.75) and (3.79) to derive

$$\begin{aligned}
II(p,z) &\leq \frac{|\det \widehat{A}_p(z) - \det A_p(z)|}{|\det \widehat{A}_p(z) \cdot \det A_p(z)|} \cdot \left\| A_p(z)_{adj} \right\| \\
&\leq \frac{1}{\delta_0^2} \cdot \left\| A_p(z)_{adj} \right\| \cdot |\det \widehat{A}_p(z) - \det A_p(z)| \\
&\leq \frac{q^2 C''}{\delta_0^2} \cdot \sup_{|z| \leq 1+(1/p)} |\det \widehat{A}_p(z) - \det A_p(z)|.
\end{aligned}$$

Since this bound is independent of z we get from (3.74) for all $p \geq p_3$

$$\max_{|z|=1+(1/p)} II(p,z) \leq \frac{q^2 C''}{\delta_0^2} \cdot \sup_{|z| \leq 1+(1/p)} |\det \widehat{A}_p(z) - \det A_p(z)|$$

$$= \frac{1}{p^2} \cdot \mathcal{O}_P(1).$$

Together with (3.77) and (3.78), this yields (3.76), which completes the proof. □

Proof of Lemma 3.11, assertion (3.25):

Let n be large enough such that $p = p(n) > p_3$, where p_3 is the constant defined in Lemma 3.8. We then have from Lemma 3.8

$$\left\| \widehat{B}_j(p) - B_j(p) \right\| \leq \Big(1 + (1/p)\Big)^{-j} p^{-2}\, \mathcal{O}_P(1)$$

uniformly for all $j \in \mathbb{N}$. Therefore,

$$\begin{aligned} \sum_{j=1}^{\infty} \left\| \widehat{B}_j(p) - B_j(p) \right\| &\leq p^{-2} \sum_{j=1}^{\infty} \Big(1 + (1/p)\Big)^{-j} \cdot \mathcal{O}_P(1) \\ &\leq p^{-2} \left(\frac{1}{1 - (1 + (1/p))^{-1}} - 1 \right) \cdot \mathcal{O}_P(1) \\ &= \mathcal{O}_P(1/p). \end{aligned} \tag{3.80}$$

Also, since $p > p_3 \geq p_2$, we can apply Lemma 3.3 and (3.9) to derive

$$\begin{aligned} \sum_{j=1}^{\infty} \left\| B_j(p) - B_j \right\| &\leq \sum_{j=1}^{\infty} (1 + |j|) \left\| B_j(p) - B_j \right\| \\ &\leq C \cdot \sum_{j=p+1}^{\infty} (1 + |j|) \left\| A_j \right\| = o(1). \end{aligned} \tag{3.81}$$

We now get

$$\sum_{j=1}^{\infty} \left\| \widehat{B}_j(p) - B_j \right\| \leq \sum_{j=1}^{\infty} \left\| \widehat{B}_j(p) - B_j(p) \right\| + \sum_{j=1}^{\infty} \left\| B_j(p) - B_j \right\| = o_P(1)$$

from (3.80) and (3.81), which completes the proof. □

Proof of Lemma 3.11, assertion (3.26):

For $p \geq p_3$ (cf. Lemma 3.8 for definition of p_3) we have

$$\begin{aligned} &\sum_{j=1}^{\infty} j \left\| \widehat{B}_j(p) \right\| \\ \leq\ &\sum_{j=1}^{\infty} j \left\| \widehat{B}_j(p) - B_j(p) \right\| + \sum_{j=1}^{\infty} j \left\| B_j(p) - B_j \right\| + \sum_{j=1}^{\infty} j \left\| B_j \right\| \end{aligned} \tag{3.82}$$

Since the inequality in Lemma 3.8 is uniform in $k \in \mathbb{N}$, we can use it to derive, for the first summand on the right-hand side of (3.82),

$$\begin{aligned}
\sum_{j=1}^{\infty} j \left\| \widehat{B}_j(p) - B_j(p) \right\| &\leq p^{-2} \sum_{j=1}^{\infty} j \left(1 + (1/p)\right)^{-j} \cdot \mathcal{O}_P(1) \\
&\leq p^{-2} \sum_{j=1}^{\infty} (j+1) \left(1 + (1/p)\right)^{-j} \cdot \mathcal{O}_P(1) \\
&\leq p^{-2} \left(\frac{1}{1 - (1 + (1/p))^{-1}} \right)^2 \cdot \mathcal{O}_P(1) \\
&= p^{-2} \left((p+1)^2 - 1 \right) \cdot \mathcal{O}_P(1) \\
&\leq 3 \cdot \mathcal{O}_P(1).
\end{aligned}$$

Also, since $p > p_3 \geq p_2$, we can bound the second and third summand on the right-hand side of (3.82) as per

$$\begin{aligned}
&\sum_{j=1}^{\infty} j \left\| B_j(p) - B_j \right\| + \sum_{j=1}^{\infty} j \left\| B_j \right\| \\
\leq\ & C \cdot \sum_{j=1}^{\infty} (1 + |j|) \left\| A_j \right\| + \sum_{j=1}^{\infty} (1 + |j|) \left\| B_j \right\| < \infty,
\end{aligned}$$

where we have used Lemma 3.3. The bound is finite due to (3.9). Combining these two bounds in (3.82) yields

$$\sum_{j=1}^{\infty} j \, \| \widehat{B}_j(p) \| \leq \mathcal{O}_P(1) \quad \text{uniformly } \ \forall \, p \geq p_3$$

□

Proof of Lemma 3.11, assertion (3.27):

Since $\underline{\varepsilon}_t^*$ is uniformly distributed on the set of centered residuals $\underline{\widehat{\varepsilon}}_{p+1}, \ldots, \underline{\widehat{\varepsilon}}_n$, cf. step (2) of the sieve bootstrap procedure, we have for the moments of its components

$$E^*\left(\varepsilon_t^*(j)^{2w}\right) = \frac{1}{n-p} \sum_{t=p+1}^{n} \widehat{\varepsilon}_t(j)^{2w}.$$

Per definition and according to representation (3.5) we also get

$$\widehat{\varepsilon}_t(j) = \varepsilon_t'(j) - \overline{\varepsilon}(j) = \varepsilon_t(j) + Q_{t,p}(j) + R_{t,p}(j) - \overline{\varepsilon}(j), \tag{3.83}$$

where

$$Q_{t,p}(j) = \sum_{k=1}^{p} (A_k(p) - \widehat{A}_k(p))^{(j,\cdot)} \, \underline{X}_{t-k} \ \text{ and } \ R_{t,p}(j) = \sum_{k=1}^{\infty} (A_k - A_k(p))^{(j,\cdot)} \, \underline{X}_{t-k}$$

and $A_k(p) := 0$ for $k > p$. Here, $A^{(j,\cdot)}$ denotes the j-th row vector of the matrix A. Since we assume in Theorem 3.10 that

$$\frac{1}{n-p}\sum_{t=p+1}^{n} \varepsilon_t(j)^{2w} \xrightarrow{P} E\Big(\varepsilon_1(j)^{2w}\Big) \qquad \forall\, w \leq h+2,$$

the desired assertion follows if we can show

$$\frac{1}{n-p}\sum_{t=p+1}^{n} \Big(\widehat{\varepsilon}_t(j)^{2w} - \varepsilon_t(j)^{2w}\Big) \xrightarrow{P} 0 \qquad \forall\, w \leq h+2. \tag{3.84}$$

Let $w \leq h+2$. Plugging (3.83) into (3.84) and using a binomial expansion (denoted in an straightforward way with $\underline{a} \in \mathbb{N}^4$ and some binomial coefficients $b_{\underline{a}}$ which can be bounded by a constant C) leads to

$$\begin{aligned}
&\left|\frac{1}{n-p}\sum_{t=p+1}^{n} \Big(\widehat{\varepsilon}_t(j)^{2w} - \varepsilon_t(j)^{2w}\Big)\right| \\
=\ &\left|\frac{1}{n-p}\sum_{t=p+1}^{n} \Big(\sum_{|\underline{a}|=2w} b_{\underline{a}}\, \varepsilon_t(j)^{a_1}\, Q_{t,p}(j)^{a_2}\, R_{t,p}(j)^{a_3}\, (-\overline{\varepsilon}(j))^{a_4} - \varepsilon_t(j)^{2w}\Big)\right| \\
\leq\ &\left|\frac{1}{n-p}\sum_{t=p+1}^{n} \sum_{|\underline{a}|=2w,\ a_1\neq 2w} b_{\underline{a}}\, \varepsilon_t(j)^{a_1}\, Q_{t,p}(j)^{a_2}\, R_{t,p}(j)^{a_3}\, (-\overline{\varepsilon}(j))^{a_4}\right| \\
\leq\ & C\cdot \sum_{|\underline{a}|=2w,\ a_1\neq 2w} \frac{1}{n-p}\sum_{t=p+1}^{n} |\varepsilon_t(j)|^{a_1}\, |Q_{t,p}(j)|^{a_2}\, |R_{t,p}(j)|^{a_3}\, |\overline{\varepsilon}(j)|^{a_4} \\
\leq\ & C\cdot \sum_{|\underline{a}|=2w,\ a_1\neq 2w} (I_n)^{a_1/2w}\, (II_n)^{a_2/2w}\, (III_n)^{a_3/2w}\, (IV_n)^{a_4/2w},
\end{aligned} \tag{3.85}$$

where Hölder's inequality was used in the final step and, moreover,

$$(I_n) = \frac{1}{n-p}\sum_{t=p+1}^{n} |\varepsilon_t(j)|^{2w},\ (II_n) = \frac{1}{n-p}\sum_{t=p+1}^{n} |Q_{t,p}(j)|^{2w},$$
$$(III_n) = \frac{1}{n-p}\sum_{t=p+1}^{n} |R_{t,p}(j)|^{2w},\ (IV_n) = \frac{1}{n-p}\sum_{t=p+1}^{n} |\overline{\varepsilon}(j)|^{2w}.$$

As the number of summands in (3.85) does not depend on n and since there is no summand with $a_2 = a_3 = a_4 = 0$, (3.84) holds true if we can show

$$(I_n) = \mathcal{O}_P(1),\ (II_n) = o_P(1),\ (III_n) = o_P(1),\ (IV_n) = o_P(1). \tag{3.86}$$

In the first step we show $(IV_n) = o_P(1)$ which is equivalent to $|\overline{\varepsilon}(j)| = o_P(1)$. From the definition of $\overline{\varepsilon}$ and from (3.83) we get

$$|\overline{\varepsilon}(j)| \leq \left|\frac{1}{n-p}\sum_{t=p+1}^{n} \varepsilon_t(j)\right| + \left|\frac{1}{n-p}\sum_{t=p+1}^{n} Q_{t,p}(j)\right| + \left|\frac{1}{n-p}\sum_{t=p+1}^{n} R_{t,p}(j)\right|. \tag{3.87}$$

The first summand on the right-hand side of (3.87) obviously converges to zero in probability because of the WLLN (remember that the random variables $\underline{\varepsilon}_t$ are uncorrelated and have mean zero). Considering the second summand we get from (3.3) for all $t \in \mathbb{Z}$

$$\begin{aligned}
\left|Q_{t,p}(j)\right| &\leq \sum_{k=1}^{p} \left|(A_k(p) - \widehat{A}_k(p))^{(j,\cdot)}\, \underline{X}_{t-k}\right| \\
&\leq \sum_{k=1}^{p}\sum_{s=1}^{q} \left|(A_k(p) - \widehat{A}_k(p))^{(j,s)}\right| \left|\underline{X}_{t-k}^{(s)}\right| \\
&\leq \sum_{k=1}^{p}\sum_{s=1}^{q} \left\|A_k(p) - \widehat{A}_k(p)\right\| \cdot \left|\underline{X}_{t-k}^{(s)}\right| \\
&\leq q^{1/2} \cdot \Big(\sum_{k=1}^{p} \left\|A_k(p) - \widehat{A}_k(p)\right\|^2\Big)^{1/2} \cdot \Big(\sum_{k=1}^{p}\sum_{s=1}^{q} \left|\underline{X}_{t-k}^{(s)}\right|^2\Big)^{1/2}, \quad (3.88)
\end{aligned}$$

where Hölder's inequality was used in the final step. Considering the third factor on the right-hand side of (3.88), we can easily see from $P\{|\underline{X}_{t-k}^{(s)}|^2 > M\} \leq E\,|\underline{X}_{t-k}^{(s)}|^2/M$ and the moment condition in Assumption 5 that $|\underline{X}_{t-k}^{(s)}|^2 = \mathcal{O}_P(1)$ uniformly for all t, k and s. Thus, the third factor of (3.88) can be bounded by $\mathcal{O}_P(p^{1/2})$. Since Assumption 6 guarantees $\sum_{k=1}^{p} \|A_k(p) - \widehat{A}_k(p)\| = \mathcal{O}_P(1/p^2)$, we also get

$$\sum_{k=1}^{p} \left\|A_k(p) - \widehat{A}_k(p)\right\|^2 = \mathcal{O}_P(1/p^2).$$

Inserting these bounds into (3.88) yields

$$\left|Q_{t,p}(j)\right| = \mathcal{O}_P(p^{-1/2}) \quad (3.89)$$

uniformly for all $t \in \mathbb{Z}$. Thus,

$$\left|\frac{1}{n-p}\sum_{t=p+1}^{n} Q_{t,p}(j)\right| = \mathcal{O}_P(p^{-1/2}). \quad (3.90)$$

Now consider the third summand on the right-hand side of (3.87). From the definition of $R_{t,p}(j)$ we get for arbitrary $t \in \mathbb{Z}$

$$|R_{t,p}(j)| \leq \sum_{k=1}^{p} \left|(A_k - A_k(p))^{(j,\cdot)}\, \underline{X}_{t-k}\right| + \sum_{k=p+1}^{\infty} \left|A_k^{(j,\cdot)}\, \underline{X}_{t-k}\right|. \quad (3.91)$$

Recall that we assume Assumption 5 to hold for $r = 1$ which delivers two useful results. Firstly, the coefficients A_k fulfil (3.9) for $r = 1$, i.e. $\sum_{k=1}^{\infty}(1 + |k|)\,\|A_k\| \leq$

$C < \infty$. Secondly, Baxter's inequality (cf. Lemma 3.1) is valid for $r = 0$, i.e. there exists $p_1 \in \mathbb{N}$ such that for all $p \geq p_1$

$$\sum_{k=1}^{p} \|A_k - A_k(p)\| \leq C' \cdot \sum_{k=p+1}^{\infty} \|A_k\|.$$

Applying these two assertions as well as Markov's inequality to (3.91), and using the uniform bound $E|\underline{X}_{t-k}^{(s)}| \leq C$ (cf. Assumption 5), yields for arbitrary $M > 0$, $t \in \mathbb{Z}$ and all $p \geq p_1$

$$\begin{aligned}
& P\Big(p \cdot |R_{t,p}(j)| > M\Big) \\
\leq\ & \frac{p}{M} \sum_{s=1}^{q} \Big(\sum_{k=1}^{p} \Big|(A_k - A_k(p))^{(j,s)}\Big| \cdot E\Big|\underline{X}_{t-k}^{(s)}\Big| + \sum_{k=p+1}^{\infty} \Big|A_k^{(j,s)}\Big| \cdot E\Big|\underline{X}_{t-k}^{(s)}\Big|\Big) \\
\leq\ & \frac{Cqp}{M} \Big(\sum_{k=1}^{p} \Big\|A_k - A_k(p)\Big\| + \sum_{k=p+1}^{\infty} \Big\|A_k\Big\|\Big) \\
\leq\ & \frac{Cqp}{M} \Big(C' \cdot \sum_{k=p+1}^{\infty} \|A_k\| + \sum_{k=p+1}^{\infty} \|A_k\|\Big) \\
\leq\ & \frac{Cq(C'+1)}{M} \sum_{k=p+1}^{\infty} (1 + |k|)\, \|A_k\| \\
\leq\ & \frac{Cq(C'+1)}{M} \sum_{k=1}^{\infty} (1 + |k|)\, \|A_k\| < \infty.
\end{aligned}$$

Note that the right-hand side does not depend on n or t and that it becomes arbitrarily small with an adequate choice of M. Therefore, we have derived

$$|R_{t,p}(j)| = \mathcal{O}_P(p^{-1}) \tag{3.92}$$

uniformly for all $t \in \mathbb{Z}$, which also gives

$$\Big|\frac{1}{n-p} \sum_{t=p+1}^{n} R_{t,p}(j)\Big| = \mathcal{O}_P(p^{-1}).$$

Combining this with (3.87) and (3.90) gives $(IV_n) = o_P(1)$.

We now turn our attention to expression (II_n). In (3.89) we have already shown $|Q_{t,p}(j)| = \mathcal{O}_P(p^{-1/2})$ uniformly for all $t \in \mathbb{Z}$. This immediately yields for all $1 \leq w \leq h + 2$

$$\frac{1}{n-p} \sum_{t=p+1}^{n} |Q_{t,p}(j)|^{2w} = \mathcal{O}_P(p^{-w}),$$

i.e. $(II_n) = o_P(1)$. Analogously, since $|R_{t,p}(j)| = \mathcal{O}_P(p^{-1})$ from (3.92), we have for all $1 \leq w \leq h + 2$

$$\frac{1}{n-p} \sum_{t=p+1}^{n} |R_{t,p}(j)|^{2w} = \mathcal{O}_P(p^{-2w}),$$

i.e. $(III_n) = o_P(1)$. Since we only look at even exponents, we get for (I_n)

$$\frac{1}{n-p}\sum_{t=p+1}^{n}|\varepsilon_t(j)|^{2w} = \frac{1}{n-p}\sum_{t=p+1}^{n}\varepsilon_t(j)^{2w} \xrightarrow{P} E\Big(\varepsilon_1(j)^{2w}\Big)$$

from the assumptions in Theorem 3.10. Hence, the moment condition from Theorem 3.10 yields

$$\frac{1}{n-p}\sum_{t=p+1}^{n}|\varepsilon_t(j)|^{2w} = \mathcal{O}_P(1),$$

which delivers the final assertion of (3.86) and completes the proof. □

Proof of Lemma 3.11, assertion (3.28):
$(\underline{\varepsilon}_t)$ is strictly stationary and we denote the distribution function of $\underline{\varepsilon}_t$ by F and the distribution itself by P_F. Given a data sample $\underline{X}_1,\ldots,\underline{X}_n$, the bootstrap innovations $\underline{\varepsilon}_t^*$ have the distribution function $\widehat{F}_n$ which is the empirical distribution function of the centered residuals $\{\widehat{\underline{\varepsilon}}_{p+1},\ldots,\widehat{\underline{\varepsilon}}_n\}$, cf. the definition of the sieve bootstrap procedure. Denote the empirical distribution function of $\{\underline{\varepsilon}_{p+1},\ldots,\underline{\varepsilon}_n\}$ by F_n and the corresponding distribution by P_{F_n}. Note that for any $\underline{x}\in\mathbb{R}^q$ both $\widehat{F}_n(\underline{x})$ and $F_n(\underline{x})$ are random variables on the same probability space while F is a deterministic function. Therefore, the Mallows metric $d_2(\widehat{F}_n, F)$ itself is a random variable. In the following we show $d_2(\widehat{F}_n, F)\to 0$ in P-probability which yields (3.28).
Since the Mallows metric fulfils

$$d_2(\widehat{F}_n, F) \;\le\; d_2(\widehat{F}_n, F_n) + d_2(F_n, F), \tag{3.93}$$

we can treat the two terms on the right-hand side seperately. For the second one we use Lemma 8.3 (a) in Bickel and Freedman (1981). The required convergence of moments can be ensured by

$$\begin{aligned}
&\left|\int \|\underline{x}\|_e^2\ P_{F_n}(d\underline{x}) - \int \|\underline{x}\|_e^2\ P_F(d\underline{x})\right| \\
=\ &\left|\frac{1}{n-p}\sum_{t=p+1}^{n}\|\underline{\varepsilon}_t\|_e^2 - E\,\|\underline{\varepsilon}_t\|_e^2\right| \\
\le\ &\sum_{j=1}^{q}\left|\frac{1}{n-p}\sum_{t=p+1}^{n}\Big(\varepsilon_t(j)^2 - E\Big(\varepsilon_t(j)^2\Big)\Big)\right| = o_P(1),
\end{aligned}$$

where $\|\cdot\|_e$ denotes the Euclidean norm and we have used Assumption 8 (for $w=1$) in the final step. This, together with Assumption 8, yields $d_2(F_n, F)\longrightarrow 0$ in P-probability via Lemma 8.3 in Bickel and Freedman (1981).

As in Bühlmann (1997), Lemma 5.4, we define for fixed n the random variable S (on a different probability space than the one that supports $d_2(\widehat{F}_n, F_n)$) to be uniformly distributed on $\{p+1, \ldots, n\}$. $\underline{Z}_1 := \underline{\widehat{\varepsilon}}_S$ and $\underline{Z}_2 := \underline{\varepsilon}_S$ then have distribution functions $\widehat{F}_n$ and F_n, respectively. The definition of the Mallows metric gives

$$\begin{aligned}
d_2(\widehat{F}_n, F_n)^2 &\leq E_S \, \|\underline{Z}_1 - \underline{Z}_2\|_e^2 \\
&= \frac{1}{n-p} \sum_{t=p+1}^{n} \|\underline{\widehat{\varepsilon}}_t - \underline{\varepsilon}_t\|_e^2 \\
&= \sum_{j=1}^{q} \frac{1}{n-p} \sum_{t=p+1}^{n} \left(\underline{\widehat{\varepsilon}}_t(j) - \underline{\varepsilon}_t(j)\right)^2 \\
&= \sum_{j=1}^{q} \frac{1}{n-p} \sum_{t=p+1}^{n} \left(Q_{t,p}(j) + R_{t,p}(j) - \underline{\overline{\varepsilon}}(j)\right)^2, \qquad (3.94)
\end{aligned}$$

where the last equation comes from (3.83). Using (3.86) one can easily obtain from (3.94) that $d_2(\widehat{F}_n, F_n) \longrightarrow 0$ in P-probability. Therefore, we get from (3.93) $d_2(\widehat{F}_n, F) \longrightarrow 0$ in P-probability which yields (3.28). $\Box$

Proof of Lemma 3.11, assertion (3.29):
Due to the Cramér-Wold device, (3.29) is equivalent to

$$\sum_{i=1}^{d} \sum_{j=1}^{q} c_{i,j} X^*_{t_i}(j) \xrightarrow{d^*} \sum_{i=1}^{d} \sum_{j=1}^{q} c_{i,j} \widetilde{X}_{t_i}(j) \ \text{ in } P\text{-prob. } \forall\, t_1, \ldots, t_d \in \mathbb{Z},\ \forall c_{i,j} \in \mathbb{R},$$

where $X^*_{t_i}(j)$ denotes the j-th component of the vector $\underline{X}^*_{t_i}$. Of course, this follows immediately if we can show

$$X^*_t(j) \xrightarrow{d^*} \widetilde{X}_t(j) \ \text{ in } P\text{-prob.} \qquad (3.95)$$

This assertion can be shown almost exactly along the lines of Lemma 5.5 in Bühlmann (1997) with only minor adaptations to our setting to consider. In the following we will therefore only show how to adapt Bühlmann's proof, using his notation as far as possible. Firstly, having (3.22) in mind and defining $\widehat{B}_0(p) := I$ for abbreviation, each component of our bootstrap process can be decomposed as

$$X^*_t(j) = \sum_{s=1}^{q} \sum_{k=0}^{\infty} \widehat{B}_k(p)^{(j,s)} \, \varepsilon^*_{t-k}(s) = \sum_{s=1}^{q} \sum_{k=0}^{M} B_k^{(j,s)} \, \varepsilon^*_{t-k}(s) + U^*_{t,n}(j) + V^*_{t,n}(j),$$

where

$$U^*_{t,n}(j) := \sum_{s=1}^{q} \sum_{k=0}^{M} \left(\widehat{B}_k(p)^{(j,s)} - B_k^{(j,s)}\right) \varepsilon^*_{t-k}(s)\,,\ V^*_{t,n}(j) := \sum_{s=1}^{q} \sum_{k=M+1}^{\infty} \widehat{B}_k(p)^{(j,s)} \, \varepsilon^*_{t-k}(s).$$

The bootstrap innovation process $(\underline{\varepsilon}_t^*)$ is (conditionally) i.i.d. Thus, we get for arbitrary $\gamma, \kappa > 0$, using (3.3),

$$\begin{aligned} P^*\Big(|U_{t,n}^*(j)| > \gamma/2\Big) &\leq (2/\gamma)\, E^* \Big| \sum_{s=1}^{q} \sum_{k=0}^{M} \Big(\widehat{B}_k(p)^{(j,s)} - B_k^{(j,s)}\Big)\, \varepsilon_{t-k}^*(s) \Big| \\ &\leq (2/\gamma) \sum_{s=1}^{q} E^*|\varepsilon_t^*(s)| \sum_{k=0}^{M} \Big|\widehat{B}_k(p)^{(j,s)} - B_k^{(j,s)}\Big| \\ &\leq (2/\gamma) \sum_{s=1}^{q} E^*|\varepsilon_t^*(s)| \sum_{k=0}^{\infty} \Big\| \widehat{B}_k(p) - B_k \Big\| \end{aligned}$$

But since (3.27) yields $E^*|\varepsilon_t^*(s)| = \mathcal{O}_P(1)$ and (3.25) delivers $\sum_{k=0}^{\infty} \Big\| \widehat{B}_k(p) - B_k \Big\| = o_P(1)$, we can choose M such that, for n sufficiently large,

$$P^*\Big(|U_{t,n}^*(j)| > \gamma/2\Big) \leq \kappa/2 \;\text{ in } P\text{-probability.} \tag{3.96}$$

Analogously, we get, for n sufficiently large,

$$P^*\Big(|V_{t,n}^*(j)| > \gamma/2\Big) \leq \kappa/2 \;\text{ in } P\text{-probability} \tag{3.97}$$

from $\sum_{k=M+1}^{\infty} \|\widehat{B}_k(p)\| = \mathcal{O}_P(1/M)$, which holds true because (3.26) yields

$$M \cdot \sum_{k=M+1}^{\infty} \|\widehat{B}_k(p)\| \leq \sum_{k=M+1}^{\infty} k\, \|\widehat{B}_k(p)\| \leq \sum_{k=1}^{\infty} k\, \|\widehat{B}_k(p)\| = \mathcal{O}_P(1).$$

Now recall that the innovations of the companion process fulfil $\mathcal{L}(\underline{\widetilde{\varepsilon}}_t) = \mathcal{L}(\underline{\varepsilon}_t)$ and $(\underline{\widetilde{\varepsilon}}_t)$ is an i.i.d. process. Thus, (3.28) yields

$$\underline{\varepsilon}_t^* \xrightarrow{d^*} \underline{\widetilde{\varepsilon}}_t \;\text{ in } P\text{-prob. } \forall\, t \in \mathbb{Z},$$

which means

$$P^*\Big(\sum_{s=1}^{q} \sum_{k=0}^{M} B_k^{(j,s)}\, \varepsilon_{t-k}^*(s) \leq x \pm \gamma\Big) \xrightarrow{P} P\Big(\sum_{s=1}^{q} \sum_{k=0}^{M} B_k^{(j,s)}\, \widetilde{\varepsilon}_{t-k}(s) \leq x \pm \gamma\Big). \tag{3.98}$$

We conclude by decomposing the components of the companion process as

$$\widetilde{X}_t(j) = \sum_{s=1}^{q} \sum_{k=0}^{\infty} B_k^{(j,s)}\, \widetilde{\varepsilon}_{t-k}(s) = \sum_{s=1}^{q} \sum_{k=0}^{M} B_k^{(j,s)}\, \widetilde{\varepsilon}_{t-k}(s) + V_t(j)$$

with $V_t(j) := \sum_{s=1}^{q} \sum_{k=M+1}^{\infty} B_k^{(j,s)}\, \widetilde{\varepsilon}_{t-k}(s)$. Analogously as before, we can show

$$P\Big(|V_t(j)| > \gamma/2\Big) \leq \kappa/2$$

for n large enough using (3.9). This, together with (3.96), (3.97) and (3.98), establishes all crucial assertions used in the proof of Lemma 5.5 in Bühlmann (1997). Following along the lines of this proof yields (3.29). □

Proof of Lemma Lemma 3.16, assertion (3.32):
From the construction of the bootstrap innovations $\underline{\varepsilon}_t^*$ in Section 3.2 we get

$$\begin{aligned} E^*(\varepsilon_0^*(r)\,\varepsilon_0^*(s)) &= (n-p)^{-1}\sum_{t=p+1}^{n}\widehat{\varepsilon}_t(r)\widehat{\varepsilon}_t(s) \\ &= (n-p)^{-1}\sum_{t=p+1}^{n}\Big(\varepsilon_t(r)+\widehat{\varepsilon}_t(r)-\varepsilon_t(r)\Big)\Big(\varepsilon_t(s)+\widehat{\varepsilon}_t(s)-\varepsilon_t(s)\Big) \\ &= (n-p)^{-1}\sum_{t=p+1}^{n}\varepsilon_t(r)\,\varepsilon_t(s)+(a)+(b)+(c), \end{aligned} \tag{3.99}$$

where $(a)=(n-p)^{-1}\sum_{t=p+1}^{n}\varepsilon_t(r)\Big(\widehat{\varepsilon}_t(s)-\varepsilon_t(s)\Big)$, $(b)=(n-p)^{-1}\sum_{t=p+1}^{n}\varepsilon_t(s)\Big(\widehat{\varepsilon}_t(r)-\varepsilon_t(r)\Big)$ and $(c)=(n-p)^{-1}\sum_{t=p+1}^{n}\Big(\widehat{\varepsilon}_t(r)-\varepsilon_t(r)\Big)\Big(\widehat{\varepsilon}_t(s)-\varepsilon_t(s)\Big)$. Considering the decomposition of $\widehat{\varepsilon}_t(s)$ as in (3.83) and using Hölder's inequality yields

$$\begin{aligned} |\,(a)\,| &\leq \Big((n-p)^{-1}\sum_{t=p+1}^{n}\varepsilon_t(r)^2\cdot(n-p)^{-1}\sum_{t=p+1}^{n}\Big(\widehat{\varepsilon}_t(s)-\varepsilon_t(s)\Big)^2\Big)^{1/2} \\ &= \Big((n-p)^{-1}\sum_{t=p+1}^{n}\varepsilon_t(r)^2\cdot(n-p)^{-1}\sum_{t=p+1}^{n}\Big(Q_{t,p}(s)+R_{t,p}(s)-\overline{\varepsilon}(s)\Big)^2\Big)^{1/2} \\ &= \mathcal{O}_P(1)\cdot o_P(1)=o_P(1), \end{aligned}$$

which can easily be seen from (3.86) and from the fact that the variance of $\varepsilon_t(r)$ is uniformly bounded in t per Assumption 10. With the same arguments we also get $(b)=o_P(1)$ and $(c)=o_P(1)$. From (3.99) and again using Assumption 10 we get

$$\begin{aligned} E^*(\varepsilon_0^*(r)\,\varepsilon_0^*(s)) &= (n-p)^{-1}\sum_{t=p+1}^{n}\varepsilon_t(r)\,\varepsilon_t(s)+o_P(1) \\ &= E(\varepsilon_0(r)\,\varepsilon_0(s))+o_P(1). \end{aligned}$$

This completes the proof. □

Proof of Lemma Lemma 3.16, assertion (3.33):
From the definitions of $\underline{X}_{t,M}$ and $\underline{X}_{t,M}^*$ in the proof of Theorem 3.15 and the respective white noise properties of $(\underline{\varepsilon}_t^*)$ and $(\underline{\varepsilon}_t)$ it follows immediately

$$\mathrm{Cov}^*(\underline{X}_{0,M}^*(r),\underline{X}_{h,M}^*(s))=0=\mathrm{Cov}(\underline{X}_{0,M}(r),\underline{X}_{h,M}(s))\quad\forall\;|h|>M. \tag{3.100}$$

On the other hand, using again the notation $A^{(r,\cdot)}$ for the r-th row vector of a matrix A, we get for all $|h| \leq M$

$$\begin{aligned}
&\operatorname{Cov}^*(\underline{X}^*_{0,M}(r), \underline{X}^*_{h,M}(s)) \\
=\;& E^*\Big(\sum_{j=0}^{M} \widehat{B}_j(p)^{(r,\cdot)}\, \underline{\varepsilon}^*_{-j} \cdot \sum_{k=0}^{M} \widehat{B}_k(p)^{(s,\cdot)}\, \underline{\varepsilon}^*_{h-k} \Big) \\
=\;& \begin{cases} \sum_{j=0}^{M-|h|} \widehat{B}_j(p)^{(r,\cdot)}\, E^*(\underline{\varepsilon}^*_0\, \underline{\varepsilon}^{*T}_0)\, \widehat{B}_{j+|h|}(p)^{(s,\cdot)^T}, & h \geq 0 \\ \sum_{j=0}^{M-|h|} \widehat{B}_{j+|h|}(p)^{(r,\cdot)}\, E^*(\underline{\varepsilon}^*_0\, \underline{\varepsilon}^{*T}_0)\, \widehat{B}_j(p)^{(s,\cdot)^T}, & h < 0 \end{cases}
\end{aligned}$$

Due to the symmetry of the expression we only look at $h \geq 0$ and get

$$\begin{aligned}
&\operatorname{Cov}^*(\underline{X}^*_{0,M}(r), \underline{X}^*_{h,M}(s)) \\
=\;& \sum_{j=0}^{M-|h|} \widehat{B}_j(p)^{(r,\cdot)}\, E(\underline{\varepsilon}_0\, \underline{\varepsilon}_0{}^T)\, \widehat{B}_{j+|h|}(p)^{(s,\cdot)^T} \\
&+ \sum_{j=0}^{M-|h|} \sum_{t,v=1}^{q} \widehat{B}_j(p)^{(r,t)} \Big(E^*(\varepsilon^*_0(t)\, \varepsilon^*_0(v)) - E(\varepsilon_0(t)\, \varepsilon_0(v)) \Big)\, \widehat{B}_{j+|h|}(p)^{(s,v)}.
\end{aligned}$$

Using (3.3), the absolute value of the second summand can be bounded by

$$\sum_{j=0}^{M-|h|} \sum_{t,v=1}^{q} \|\widehat{B}_j(p)\| \cdot \Big| E^*(\varepsilon^*_0(t)\, \varepsilon^*_0(v)) - E(\varepsilon_0(t)\, \varepsilon_0(v)) \Big| \cdot \|\widehat{B}_{j+|h|}(p)\|,$$

which converges to zero in probability because of (3.32) and since $\|\widehat{B}_j(p)\| = \mathcal{O}_P(1)$ uniformly, cf. (3.26). Therefore,

$$\begin{aligned}
&\operatorname{Cov}^*(\underline{X}^*_{0,M}(r), \underline{X}^*_{h,M}(s)) \\
=\;& \sum_{j=0}^{M-|h|} \widehat{B}_j(p)^{(r,\cdot)}\, E(\underline{\varepsilon}_0\, \underline{\varepsilon}_0{}^T)\, \widehat{B}_{j+|h|}(p)^{(s,\cdot)^T} + o_P(1) \\
=\;& \sum_{j=0}^{M-|h|} B_j^{(r,\cdot)}\, E(\underline{\varepsilon}_0\, \underline{\varepsilon}_0{}^T)\, B_{j+|h|}{}^{(s,\cdot)^T} \\
&+ \sum_{j=0}^{M-|h|} \Big(\widehat{B}_j(p) - B_j \Big)^{(r,\cdot)} E(\underline{\varepsilon}_0\, \underline{\varepsilon}_0{}^T)\, \widehat{B}_{j+|h|}(p)^{(s,\cdot)^T} \\
&+ \sum_{j=0}^{M-|h|} B_j^{(r,\cdot)}\, E(\underline{\varepsilon}_0\, \underline{\varepsilon}_0{}^T) \Big(\widehat{B}_{j+|h|}(p) - B_{j+|h|} \Big)^{(s,\cdot)^T} + o_P(1).
\end{aligned}$$

Again, the absolute value of the second summand on the right-hand side can be bounded by

$$\sum_{j=0}^{M-|h|} \|\widehat{B}_j(p) - B_j\| \cdot \|E(\underline{\varepsilon}_0 \, \underline{\varepsilon}_0{}^T)\| \cdot \|\widehat{B}_{j+|h|}(p)\|,$$

which converges to zero in probability because of (3.25) from Lemma 3.11 and since $\|\widehat{B}_j(p)\| = \mathcal{O}_P(1)$ uniformly, cf. (3.26). The same holds true for the third summand on the right-hand side which yields for $0 \le h \le M$

$$\begin{aligned}
& \mathrm{Cov}^*(\underline{X}^*_{0,M}(r), \underline{X}^*_{h,M}(s)) \\
= & \sum_{j=0}^{M-|h|} B_j^{(r,\cdot)} \, E(\underline{\varepsilon}_0 \, \underline{\varepsilon}_0{}^T) \, B_{j+|h|}{}^{(s,\cdot)^T} + o_P(1) \\
= & \mathrm{Cov}(\underline{X}_{0,M}(r), \underline{X}_{h,M}(s)) + o_P(1).
\end{aligned}$$

The same result can be obtained by analogous calculations for $-M \le h < 0$ and together with (3.100) we get

$$\mathrm{Cov}^*(\underline{X}^*_{0,M}(r), \underline{X}^*_{h,M}(s)) = \mathrm{Cov}(\underline{X}_{0,M}(r), \underline{X}_{h,M}(s)) + o_P(1) \quad \forall\, h \in \mathbb{Z}.$$

This completes the proof. $\square$

Proof of Lemma Lemma 3.16, assertion (3.34):
Firstly, considering that $\widehat{\underline{m}}(x)$ is a constant conditional on the given data, we have for arbitrary $t, v \in \mathbb{N}$

$$\mathrm{Cov}^*(\underline{Y}^*_{t,M}(r), \underline{Y}^*_{v,M}(s)) = \mathrm{Cov}^*(\underline{X}^*_{t,M}(r), \underline{X}^*_{v,M}(s)).$$

With this property, the strict stationarity of $(\underline{X}^*_{t,M})$ and the definitions of $\underline{L}^*_{n,M}(x)$ and $\widehat{\underline{m}}^*_M(x)$ we can derive

$$\begin{aligned}
& E^*(\underline{L}^*_{n,M}(x)^{(r)} \cdot \underline{L}^*_{n,M}(x)^{(s)}) \\
= & n\delta \cdot \mathrm{Cov}^*(\widehat{\underline{m}}^*_M(x)^{(r)}, \widehat{\underline{m}}^*_M(x)^{(s)}) \\
= & \frac{1}{n\delta} \sum_{t=1}^{n} \sum_{v=1}^{n} K\Big(\frac{x - t/n}{\delta}\Big) K\Big(\frac{x - v/n}{\delta}\Big) \mathrm{Cov}^*(\underline{X}^*_{t,M}(r), \underline{X}^*_{v,M}(s)) \\
= & \frac{1}{n\delta} \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} K\Big(\frac{x - t/n}{\delta}\Big) K\Big(\frac{x - (t+|h|)/n}{\delta}\Big) \mathrm{Cov}^*(\underline{X}^*_{0,M}(r), \underline{X}^*_{h,M}(s)).
\end{aligned}$$

Since $\mathrm{Cov}^*(\underline{X}^*_{0,M}(r), \underline{X}^*_{h,M}(s))$ equals zero for $|h| > M$, cf. (3.100), and is otherwise given by (3.33), the right-hand side is equal to

$$\frac{1}{n\delta} \sum_{|h| \le (n-1) \wedge M} \sum_{t=1}^{n-|h|} K_{t,x} K_{t+|h|,x} \cdot \mathrm{Cov}(\underline{X}_{0,M}(r), \underline{X}_{h,M}(s))$$

$$+\frac{1}{n\delta}\sum_{|h|\le(n-1)\wedge M}\sum_{t=1}^{n-|h|}K_{t,x}K_{t+|h|,x}\cdot o_P(1), \tag{3.101}$$

where the abbreviation $K_{t,x} := K((x-t/n)/\delta)$ is used. The first summand can be decomposed as

$$\frac{1}{n\delta}\sum_{|h|\le(n-1)\wedge M}\sum_{t=1}^{n-|h|}K^2\Big(\frac{x-t/n}{\delta}\Big)\cdot \operatorname{Cov}(\underline{X}_{0,M}(r),\underline{X}_{h,M}(s))$$
$$+\frac{1}{n\delta}\sum_{|h|\le(n-1)\wedge M}\sum_{t=1}^{n-|h|}K_{t,x}(K_{t+|h|,x}-K_{t,x})\cdot \operatorname{Cov}(\underline{X}_{0,M}(r),\underline{X}_{h,M}(s)). \tag{3.102}$$

In order to determine the limit of this expression one can easily see that, under the assumptions $\delta\to 0$ and $n\delta\to+\infty$ as $n\to\infty$, $\sum_{t=1}^{n-|h|}(n\delta)^{-1}K^2((x-t/n)/\delta)$ is a Riemann approximation which converges to $\int_{-1}^{1}K^2(u)\,du$. Applying the dominated convergence Theorem to (3.102) yields the limit

$$\int_{-1}^{1}K^2(u)\,du\cdot\sum_{h=-M}^{M}\operatorname{Cov}(\underline{X}_{0,M}(r),\underline{X}_{h,M}(s))=\Sigma_M^{(r,s)}$$

for the first summand while the second summand of (3.102) converges to zero. This can be obtained because only $2n\delta$ summands of $\sum_{t=1}^{n-|h|}(n\delta)^{-1}K_{t,x}(K_{t+|h|,x}-K_{t,x})$ are non-zero since the function K is zero outside of $[-1,1]$. Therefore, considering that K is bounded and Lipschitz, the absolute value of $\sum_{t=1}^{n-|h|}(n\delta)^{-1}K_{t,x}(K_{t+|h|,x}-K_{t,x})$ can be bounded by

$$\frac{1}{n\delta}\sum_{t=1}^{n-|h|}\Big|K\Big(\frac{x-t/n}{\delta}\Big)\Big|\cdot\Big|K\Big(\frac{x-(t+|h|)/n}{\delta}\Big)-K\Big(\frac{x-t/n}{\delta}\Big)\Big|$$
$$\le\ \frac{C}{n\delta}\cdot\sum_{t=1}^{n-|h|}\Big|K\Big(\frac{x-t/n}{\delta}\Big)\Big|\cdot\frac{|h|}{n\delta}\le\frac{C}{n\delta}\cdot 2n\delta\cdot\frac{|h|}{n\delta}=o((n\delta)^{-1}).$$

Applying the dominated convergence theorem as before shows that the second summand in (3.101) converges to zero in probability. Altogether, this yields

$$E^*(\underline{L}^*_{n,M}(x)^{(r)}\cdot\underline{L}^*_{n,M}(x)^{(s)})=\Sigma_M^{(r,s)}+o_P(1).$$

□

Proof of Lemma Lemma 3.16, assertion (3.35):
Let $r,s\in\{1,\dots,q\}$ be arbitrary. From (3.5) and the definition of the truncated process $(\underline{X}_{t,M})$ in the proof of Theorem 3.15 it is obvious that $\underline{X}_t(r)=\underline{X}_{t,M}(r)+$

$\underline{X}^+_{t,M}(r)$ where $\underline{X}^+_{t,M}(r) := \sum_{j=M+1}^{\infty} B_j^{(r,\cdot)} \underline{\varepsilon}_{t-j}$. Decomposing $\underline{X}_0(r)$ and $\underline{X}_h(s)$ in this way we get

$$\begin{aligned}
&\sum_{h=-M}^{M} \Big|\mathrm{Cov}(\underline{X}_{0,M}(r), \underline{X}_{h,M}(s)) - \mathrm{Cov}(\underline{X}_0(r), \underline{X}_h(s))\Big| \\
&\leq \sum_{h=-M}^{M} \Big|\mathrm{Cov}(\underline{X}_{0,M}(r), \underline{X}^+_{h,M}(s))\Big| + \sum_{h=-M}^{M} \Big|\mathrm{Cov}(\underline{X}^+_{0,M}(r), \underline{X}_{h,M}(s))\Big| \\
&\qquad + \sum_{h=-M}^{M} \Big|\mathrm{Cov}(\underline{X}^+_{0,M}(r), \underline{X}^+_{h,M}(s))\Big|.
\end{aligned} \tag{3.103}$$

Using the usual notation for the L_2-norm, (3.3) and the fact that $(\underline{\varepsilon}_t)$ is strictly stationary in our setting, i.e. $\|\varepsilon_t(r)\|_2 = \|\varepsilon_0(r)\|_2 \leq C$ uniformly for all components $r = 1, \dots, q$, we derive for the first summand on the right-hand side

$$\begin{aligned}
&\sum_{h=-M}^{M} \Big|\mathrm{Cov}\Big(\underline{X}_{0,M}(r), \sum_{j=M+1}^{\infty} B_j^{(s,\cdot)} \underline{\varepsilon}_{h-j}\Big)\Big| \\
&\leq \sum_{h=-M}^{M} \Big(\Big\| \underline{X}_{0,M}(r) \Big\|_2 \cdot \Big\| \sum_{j=M+1}^{\infty} B_j^{(s,\cdot)} \underline{\varepsilon}_{h-j} \Big\|_2 \Big) \\
&\leq \sum_{k=0}^{M} \sum_{u=1}^{q} |B_k^{(r,u)}| \cdot \|\varepsilon_{-k}(u)\|_2 \cdot \sum_{h=-M}^{M} \sum_{j=M+1}^{\infty} \sum_{v=1}^{q} |B_j^{(s,v)}| \cdot \|\varepsilon_{h-j}(v)\|_2 \\
&\leq C \cdot \sum_{k=0}^{M} \|B_k\| \cdot (2M+1) \sum_{j=M+1}^{\infty} \|B_j\| \\
&\leq 3C \cdot \sum_{k=0}^{\infty} \|B_k\| \cdot \sum_{j=M+1}^{\infty} M\, \|B_j\| \\
&\leq 3C \cdot \sum_{k=0}^{\infty} \|B_k\| \cdot \sum_{j=M+1}^{\infty} j\, \|B_j\|,
\end{aligned}$$

which converges to zero as $M \to +\infty$ since $\sum_{j=1}^{\infty} j\, \|B_j\| < \infty$, cf. (3.9). Analogously, the same can be shown for the other two summands on the right-hand side of (3.103), i.e.

$$\sum_{h=-M}^{M} \Big|\mathrm{Cov}(\underline{X}_{0,M}(r), \underline{X}_{h,M}(s)) - \mathrm{Cov}(\underline{X}_0(r), \underline{X}_h(s))\Big| \longrightarrow 0, \quad M \to \infty,$$

which immediately yields (3.35). □

Bibliography

Altman, N.S. (1990): Kernel Smoothing of Data with Correlated Errors. *Journal of the American Statistical Association* **85**, 749–759.

Anderson, T.W. (1971): *The Statistical Analysis of Time Series*. Wiley, New York.

Baxter, G. (1962): An Asymptotic Result for the Finite Predictor. *Math. Scand.* **10**, 137–144.

Baxter, G. (1963): A Norm Inequality for a Finite-Section Wiener-Hopf Equation. *Illinois J. Math.* **7**, 97–103.

Bhatt, S.J. and Dedania, H.V. (2003): Beurling algebra analogues of the classical theorems of Wiener and Lévy on absolutely convergent Fourier series. *Proc. Indian Acad. Sci. (Math. Sci.)* **113**, 179–182.

Bickel, P.J. and Freedman, D.A. (1981): Some Asymptotic Theory for the Bootstrap. *The Annals of Statistics* **9**, 1196–1217.

Brockwell, P.J. and Davis, R.A. (1991): *Time Series: Theory and Methods (2nd Ed.)*. Springer, New York.

Bühlmann, P. (1995): Sieve Bootstrap for Time Series. *Technical Report* **431**, Department of Statistics, University of California, Berkeley.

Bühlmann, P. (1997): Sieve Bootstrap for Time Series. *Bernoulli* **3**, 123–148.

Carlstein, E. (1986): The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *The Annals of Statistics* **14**, 1171–1179.

Cheng, R. and Pourahmadi, M. (1993): Baxter's inequality and convergence of finite predictors of multivariate stochastic processes. *Probability Theory and Related Fields* **95**, 115–124.

Cressie, N. (1993): *Statistics for Spatial Data.* Wiley, New York.

Dahlhaus, R. and Janas, D. (1996): A frequency domain bootstrap for ratio statistics in time series analysis. *The Annals of Statistics* **24**, 1934–1963.

Efron, B. (1979): Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* **7**, 1–26.

Franke, J. and Härdle, W. (1992): On bootstrapping kernel spectral estimates. *The Annals of Statistics* **20**, 121–145.

Freedman, D. (1984): On bootstrapping two-stage least squares estimates in stationary linear models. *The Annals of Statistics* **12**, 827–842.

Gröchenig, K. (2007): Weight functions in time-frequency analysis. In: *Pseudodifferential Operators: Partial Differential Equations and Time-Frequency Analysis.* Rodino, L. et al. (Eds.). *Fields Institute Comm.* **52**, 343–366.

Guyon, X. (1995): *Random Fields on a Network – Modelling, Statistics and Applications.* Springer, New York.

Hannan, E.J. and Deistler, M. (1988): *The Statistical Theory of Linear Systems.* Wiley Series in Prob. and Math. Stat.

Hannan, E.J. and Kavalieris, L. (1986): Regression, Autoregression Models. *Journal of Time Series Analysis* **7**, 27–49.

Hart, J.D. (1991): Kernel Regression Estimation with Time Series Errors. *Journal of the Royal Statistical Society, Series B* **53**, 173–187.

Helson, H. and Lowdenslager, D. (1958): Prediction Theory and Fourier Series in Several Variables. *Acta Math.* **99**, 165–202.

Kreiss, J.-P. (1988): *Asymptotical Inference for a Class of Stochastic Processes.* Habilitationsschrift, Universität Hamburg.

Kreiss, J.-P. (1992): Bootstrap Procedures for AR(∞) Processes. In: *Bootstrapping and Related Techniques.* Jöckel, K.H., Rothe, G. and Sender, W. (Eds.). Lecture Notes in Economics and Mathematical Systems **376** Springer, Heidelberg, 107–113.

Kreiss, J. P. (1997): Asymptotic properties of residual bootstrap for autoregressions. Mimeo, Institute of Mathematical Stochastics, Technical University of Braunschweig, Germany.

Kreiss, J.-P., and Paparoditis, E. (2003): Autoregressive aided periodogram bootstrap. *The Annals of Statistics* **31**, 1923–1955.

Kreiss, J.-P., Paparoditis, E. and Politis, D. (2011): On the Range of Validity of the Autoregressive Sieve Bootstrap. *The Annals of Statistics* **39**, 2103–2130.

Künsch, H.R. (1989): The Jackknife and the Bootstrap for General Stationary Observations. *The Annals of Statistics* **17**, 1217–1241.

Lütkepohl, H. (1996): *Handbook of Matrices*. Wiley, Chichester.

Paparoditis, E. and Streitberg, B. (1991): Order Identification Statistics in Stationary Autoregressive Moving Average Models: Vector Autocorrelations and the Bootstrap. *Journal of Time Series Analysis* **13**, 415–435.

Paparoditis, E. (1996): Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes. *Journal of Multivariate Analysis* **57**, 277–296.

Pourahmadi, M. (2001): *Foundations of Time Series Analysis and Prediction Theory*. Wiley, 2001.

Romano, J. P. and Wolf, M. (2000): A more general central limit theorem for m-dependent random variables with unbounded m. *Statistics and Probability Letters* **47**, 115–124.

Whittle, P. (1954): On Stationary Processes in the Plane. *Biometrika* **41**, 434–449.

Whittle, P. (1963): On the Fitting of Multivariate Autoregressions, and the Approximate Canonical Factorization of a Spectral Density Matrix. *Biometrika* **50**, 129–134.

Wiener, N. and Masani, P. (1957): The Prediction Theory of Multivariate Stochastic Processes, I. *Acta Mathematica* **98**, 111–150.

Wiener, N. and Masani, P. (1958): The Prediction Theory of Multivariate Stochastic Processes, II. *Acta Mathematica* **99**, 93–137.

Yokoyama, R. (1980): Moment Bounds for Stationary Mixing Sequences. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **52**, 45–57.

Zygmund, A. (2002): *Trigonometric Series – Third Edition*. Cambridge University Press, Cambridge.

Curriculum Vitae

Personal data

Name	Marco Meyer
born on	05/19/1983
in	Peine

Professional experience and education

07/2010 - current	Research and teaching assistant at the Institute of Mathematical Stochastics, Technical University of Braunschweig
10/2013 - 03/2014	partly funded by RTG 1953 ”Statistical Modelling of Complex Systems and Processes” of the German Research Foundation
10/2003 - 06/2010	Studies of Financial and Business Mathematics at Technical University of Braunschweig (degree: diploma in Financial and Business Mathematics)
07/2002 - 04/2003	Alternative civilian service at Arbeiter-Samariter-Bund Peine e.V.
06/2002	Abitur (general university admission) at Gymnasium am Silberkamp in Peine

Lebenslauf

Persönliche Daten

Name	Marco Meyer
geboren am	19.05.1983
in	Peine

Werdegang

07/2010 - dato	Wissenschaftlicher Mitarbeiter am Institut für Mathematische Stochastik der Technischen Universität Braunschweig
10/2013 - 03/2014	teilfinanziert durch GRK 1953 "Statistical Modelling of Complex Systems and Processes" der Deutschen Forschungsgemeinschaft
10/2003 - 06/2010	Studium der Finanz- und Wirtschaftsmathematik an der Technischen Universität Braunschweig (Abschluss: Diplom-Finanz- und Wirtschaftsmathematiker)
07/2002 - 04/2003	Ableistung des Zivildienstes beim Arbeiter-Samariter-Bund Peine e.V.
06/2002	Abitur am Gymnasium am Silberkamp in Peine